Greed is Super: A Fast Algorithm for Super-Resolution

Greed is Super: A Fast Algorithm for Super-Resolution

Armin Eftekhari and Michael B. Wakin*

November 12, 2015

Abstract

We present a fast two-phase algorithm for super-resolution with strong theoretical guarantees. Given the low-frequency part of the spectrum of a sequence of impulses, Phase I consists of a greedy algorithm that roughly estimates the impulse positions. These estimates are then refined by local optimization in Phase II.

In contrast to the convex relaxation proposed by Candès et al., our approach has a low computational complexity but requires the impulses to be separated by an additional logarithmic factor to succeed. The backbone of our work is the fundamental work of Slepian et al. involving discrete prolate spheroidal wave functions and their unique properties.

Keywords— Super-resolution, Parameter estimation, Greedy algorithms, Local optimization, Discrete prolate spheroidal wave functions, Slepian functions
AMS Subject Classifications— 94A12, 94A15, 42A99

1 Introduction

Many sensing mechanisms have finite resolution or bandwidth. Provided with the low-frequency content of the signal, *super-resolution* is then the problem of (partially or completely) recovering the high-frequency content of the signal. More concretely, here we restrict ourselves to the problem set up next.

Consider the time interval $\mathbb{I} = [0, 1)$. For integer K, $\tau \in \mathbb{R}^K$, and $\alpha \in \mathbb{R}^K$—all unknown—consider the signal $x_{\tau,\alpha}(t) = \sum_{i=1}^{K} \alpha[i] \cdot \delta(t \ominus \tau[i])$ where $\delta(\cdot)$ is the Dirac delta function and $\ominus$ denotes subtraction with wraparound on $\mathbb{I}$.[1] The signal $x_{\tau,\alpha}(\cdot)$ can be characterized by its Fourier series coefficients $\{\widehat{x}_{\tau,\alpha}[l]\}_l$, where

$$\widehat{x}_{\tau,\alpha}[l] = \left\langle x_{\tau,\alpha}(t), e^{\mathrm{i}2\pi l t} \right\rangle_{\mathbb{I}}, \qquad l \in \mathbb{Z}.$$

For a cut-off frequency $f_C \in \mathbb{N}$, we wish to recover K, τ, and α from the low-frequency content of $x_{\tau,\alpha}(\cdot)$, namely the coefficients $\{\widehat{x}_{\tau,\alpha}[l]\}$, $|l| \le f_C$. Equivalently, through an ideal low-pass filter with cut-off frequency f_C, we observe $y(t) := \sum_{i=1}^{K} \alpha[i] \cdot D_{f_C}(t \ominus \tau[i])$ and wish to recover the unknowns. Here, $D_{f_C}(\cdot)$ is the Dirichlet kernel[2] of width approximately $1/f_C$ in time.

*Both authors are with the Department of Electrical Engineering and Computer Science at the Colorado School of Mines. The authors' email addresses are armin.eftekhari@gmail.com and mwakin@mines.edu. AE is the corresponding author. This work was partially supported by NSF CAREER Grant CCF-1149225 and NSF Grant CCF-1409261.

[1] Later on, we will slightly modify the notation in the interest of mathematical rigor.

[2] The Dirichlet kernel is sometimes referred to as the "digital" sinc.

1.1 Our Approach

We focus on estimating the positions τ, since an estimate of the amplitudes α can subsequently be obtained using least-squares. When $K = 1$, the matched filter (e.g., [9]) provides the optimal solution to the problem. Our approach is to generalize the matched filter as follows.

We propose to iteratively find the largest peak of the measured signal $y(\cdot)$ and, in order to avoid falsely detecting nearby points in subsequent iterations, erase the neighborhood of each peak. Unfortunately, because of the heavy tail and slow decay of the Dirichlet kernel, this approach is only effective when the impulses are widely separated, the noise is negligible, and the dynamic range $\max_i |\alpha[i]| \,/\, \min_i |\alpha[i]|$ is small.

To overcome this setback, we first *filter*[3] the measurement signal $y(\cdot)$ with a *kernel* $g_{\sigma,f_C}(\cdot)$ that is band-limited to $[-f_C, f_C]$ in frequency and decays rapidly outside of the (typically small) interval $[-\sigma, \sigma]$ in time. More specifically, set $N = 2f_C + 1$ for short. After setting $\sigma = \frac{c}{N}$ for a factor c, our approach is to first filter $y(\cdot)$ with $g_{\sigma,N}(\cdot)$ and then iteratively select the peaks of the output of the filter, while removing the neighborhood of each peak to avoid false detections (as outlined in the previous paragraph).

The obtained estimate of the position vector τ can then be refined by posing super-resolution as a non-convex program, which we solve (using the projected Newton's method) with the output of the greedy search above as the initial point.

For the choice of kernel $g_{\sigma,N}(\cdot)$, we recommend the top *discrete prolate spheroidal wave function* (DPSWF) [24].[4] Given $\sigma \in (0, \frac{1}{2})$ and integer $N = 2f_C + 1$, the top DPSWF $\psi_{0,\sigma,N}(\cdot)$ is optimal in that, among all signals supported on $\mathbb{I}$ in time and $[-f_C, f_C]$ in frequency, $\psi_{0,\sigma,N}(\cdot)$ is maximally concentrated (in L_2 sense) on the small interval $[0, \sigma] \cup [1 - \sigma, 1)$ in time (see Figure 1(a)).

The resulting "two-phase" algorithm is very fast, in part because fast and convenient means for generating DPSWFs exist [22]. Moreover, in the absence of noise, this algorithm exactly recovers the impulse positions. As the noise level increases, the quality of the output gradually deteriorates. We will thoroughly verify these claims in later sections.

As an example, let the cut-off frequency $f_C = 50$, and set

$$\tau = [0.2995\ 0.3663\ 0.4332\ 0.5000\ 0.5668\ 0.6337\ 0.7005]^T, \qquad \text{(positions)}$$

$$\alpha = [10\ -1\ 1\ -3\ 2\ -5\ 2]^T. \qquad \text{(amplitudes)}$$

The measured (low-frequency) signal $y(\cdot)$ is depicted in Figure 1(b). Note that the impulses are separated by roughly only $3/f_C$. We then set $\sigma = \frac{3/2}{2f_C+1} = 0.0149$ for the top DPSWF $\psi_{0,\sigma,N}(\cdot)$. In this case, the greedy step produces an estimate $\widehat{\tau}$ which satisfies $\|\widehat{\tau} - \tau\|_\infty \leq 0.001$. This estimate is then refined via Newton's method to recover τ perfectly up to machine precision.

[3] Filtering a signal $a(\cdot)$ with another signal $b(\cdot)$ (both in $L_2(\mathbb{I})$) produces their *circular convolution* $[a \circledast b](\cdot) = \int_{\mathbb{I}} a(t) \cdot b(\cdot \ominus t)\, dt \in L_1(\mathbb{I})$.

[4] DPSWFs are also known as the "Slepian functions" in honor of David S. Slepian.

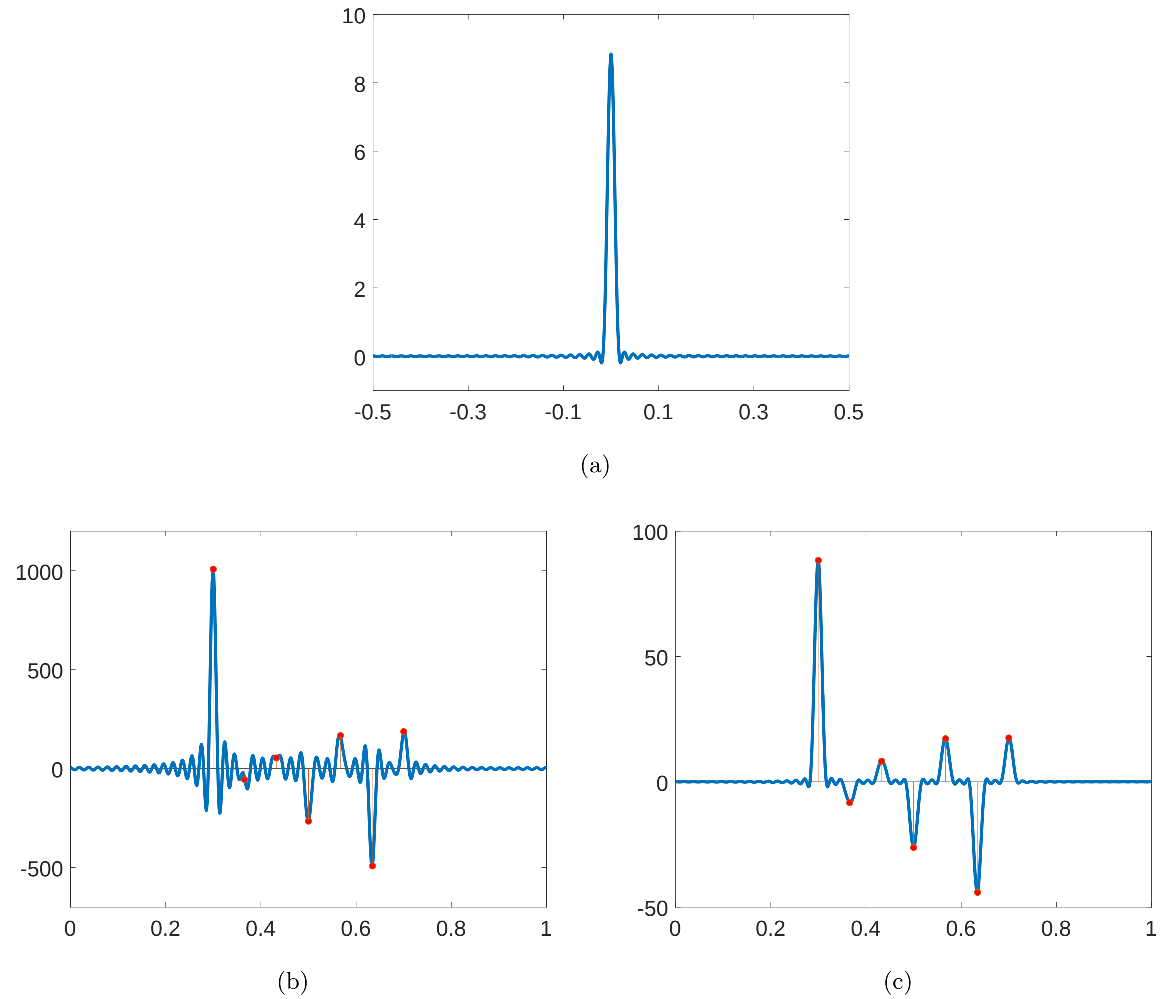

Figure 1: (a) Graph of the top DPSWF $\psi_{0,\sigma,N}(\cdot)$ versus time for bandwidth $f_C = 50$ and *effective* duration of approximately $2\sigma = \frac{3}{2f_C+1}$ in time. Note the sharp decay away from the origin. (For clarity, the domain here is $[-1/2, 1/2)$ instead of $[0, 1)$ in the text.) (b) An example of a low-resolution signal (blue) and original impulse positions (red). (c) Signal in part (b) filtered by $\psi_{0,\sigma,f_C}(\cdot)$. Note that the peaks provide a good estimate of the unknown impulse positions. The two-phase algorithm in this paper builds on this insight to return the precise location of impulses. The horizontal axis in all graphs show the time-domain.

There are different ways in which this super-resolution algorithm may be generalized. An extension to higher dimensions is of interest in, say, image processing, and replacing the Dirac delta function with a general template establishes a connection with the broad existing literature on *deconvolution* [21].

1.2 Organization

This paper is organized as follows. Section 2 gives a formal statement of the problem and collects the notation. The two-phase algorithm for super-resolution is developed in Sections 3 and 4. Phase I consists of a greedy algorithm that initializes the local optimization in Phase II. The final product is presented in Algorithms I and II (on pages 8 and 14, respectively) and is accessible even without reading the rest of the paper. The MATLAB code for the two-phase algorithm is also available online.[5]

Following the publication of [3], a steady stream of good research has gradually enriched our knowledge of this topic. Among others, [3] was followed by [26, 8, 1, 14, 7, 19, 12]. A brief survey and comparison is presented in Section 5, which is by no means exhaustive. We remark that an excerpt of this work previously appeared in [10].

The theoretical guarantees for our algorithm consist of Proposition 2 for Phase I and Theorem 6 for Phase II (and are proved in Sections 6.1 and 6.2, respectively). Our results are asymptotic and hold as $f_C \to \infty$. Naturally, these results rely heavily on certain asymptotic ($N = 2f_C + 1 \to \infty$) properties of the kernel $g_{\sigma,N}(\cdot)$ (that we identify and collect in Criteria 1, 4, and 5). The decision to opt for asymptotic guarantees was driven by the asymptotic nature of the existing machinery to study the properties of DPSWFs (the recommended kernel here). Indeed, based on empirical observations and preliminary analysis, we conjecture that DPSWFs satisfy these asymptotic criteria; formally proving this conjecture remains a topic of ongoing work.[6] Finally, despite the asymptotic nature of our results, the proposed two-phase algorithm is successful in simulations with cut-off frequency f_C as low as 50 (as we saw in the example earlier).

Lastly, to keep this paper short, we deferred the elementary calculations to the accompanying document [11].

1.3 Contributions

In this work, we develop and analyze a two-phase algorithm to resolve impulses from low-pass frequency information. Greedy algorithms for super-resolution have already appeared in the literature [12]. However, we are convinced that the present method offers certain advantages (particularly in terms of computational complexity) that are absent from the existing literature (see Section 5).

We tend to hold a similar conviction regarding the theoretical contribution of this work. In this aspect, precedents for a two-phase approach based on a good initialization followed by local optimization have appeared in other contexts [17, 4]. Nevertheless, to the best of our knowledge, this is the first work that develops theoretical guarantees for using a second-order optimization algorithm in the second phase. Rather unfortunately, that laborious task is partially responsible for the large volume of this work.

Another aspect of this work is its use of prolate functions [15, 22, 24], and particularly the top DPSWF $\psi_{0,\sigma,N}(\cdot)$. Originally published in a series of landmark papers in 1960s and 1970s, prolate functions—designed as a highly localized basis for band-limited functions—largely influenced harmonic analysis for years that followed. As shown here and in [6], prolate functions have the potential to play an important role in a variety of problems in modern signal processing as well. Perhaps another contribution of our work is then to ignite the interest of readers in these functions and their uses.

[5] http://inside.mines.edu/∼mwakin/publications.html#software

[6] We may add that the asymptotic properties of DPSWFs are not fully understood in the particular setting studied here (where $\sigma \propto \frac{1}{f_C}$ as opposed to constant σ as in [24, 23]). We also remark that, if not DPSWFs, it remains conceivable that some other functions may satisfy these or similar criteria.

2 Problem Setup

Consider the interval $\mathbb{I} = [0, 1)$ in time. We let $\oplus$ and $\ominus$ denote the addition and subtraction operators modulo one. For example, for $\rho_1, \rho_2 \in \mathbb{I}$,

$$d(\rho_1, \rho_2) := \min\left(\rho_1 \ominus \rho_2, \rho_2 \ominus \rho_1\right), \tag{1}$$

is the *wraparound distance* between ρ_1 and ρ_2 (see e.g., [3]). We study (atomic) measures on $\mathbb{I}$ of the form

$$x_{\tau,\alpha} = \sum_{i=1}^{K} \alpha[i] \cdot \delta_{\tau[i]}, \tag{2}$$

for integer K, vector of locations $\tau \in \mathbb{I}^K$ (with distinct entries), and vector of amplitudes $\alpha \in \mathbb{R}^K$. Here, $\delta_{\tau[i]}$ is the Dirac measure translated by $\tau[i] \in \mathbb{I}$. Upon existence, $x_{\tau,\alpha}$ can be completely characterized by its Fourier series $\widehat{x}_{\tau,\alpha}$, that is

$$x_{\tau,\alpha} = \sum_{l=-\infty}^{\infty} \widehat{x}_{\tau,\alpha}[l] \cdot E_l(\cdot),$$

with

$$\widehat{x}_{\tau,\alpha}[l] = \langle x_{\tau,\alpha}, E_l(\cdot)\rangle = \sum_{i=1}^{K} \alpha[i] \cdot e^{-\mathrm{i}\,2\pi l \tau[i]}, \qquad l \in \mathbb{Z}. \tag{3}$$

$$E_l(t) := e^{\mathrm{i}2\pi l t}, \qquad t \in \mathbb{R}.$$

Inner products everywhere are computed on $\mathbb{I}$ in this work.

Given only the (possibly noisy) low-frequency content of $x_{\tau,\alpha}$, we wish to infer the number of impulses K, positions $\tau \in \mathbb{I}^K$, and amplitudes $\alpha \in \mathbb{R}^K$. More specifically, for cut-off frequency $f_C \in \mathbb{N}$, if

$$y(\cdot) := \left[\mathcal{Q}_{\mathbb{F}}\left(x_{\tau,\alpha} + n\right)\right](\cdot) \in L_2(\mathbb{I}) \tag{4}$$

denotes the (low-frequency) measurement signal, we wish to recover the unknowns: K, τ, and α. Here, $\mathcal{Q}_{\mathbb{F}}(\cdot) : L_2(\mathbb{I}) \to L_2(\mathbb{I})$ is the ideal low-pass filter that restricts the frequency content of its input signal to $\mathbb{F} := [-f_C : f_C] = \{-f_C, -f_C + 1, \cdots, f_C\}$. Also, $n(\cdot) \in L_\infty(\mathbb{I})$ is the (real-valued) noise signal.

The low-pass measurement signal $y(\cdot)$ and noise $n(\cdot)$ may be written as

$$y(\cdot) = \sum_{l=-f_C}^{f_C} \widehat{y}[l] \cdot E_l(\cdot), \tag{5}$$

$$n(\cdot) = \sum_{l=-\infty}^{\infty} \widehat{n}[l] \cdot E_l(\cdot), \tag{6}$$

where $\widehat{y} = \{\widehat{y}[l]\}_l$ and $\widehat{n} = \{\widehat{n}[l]\}_l$ are the corresponding Fourier series. Then, (4) may equivalently be written as

$$\widehat{y}[l] = \widehat{x}_{\tau,\alpha}[l] + \widehat{n}[l], \qquad l \in \mathbb{F} = [-f_C : f_C]. \tag{7}$$

To reiterate, given $y(\cdot)$ or its nonzero Fourier series coefficients $\{\widehat{y}[l]\}_{l\in\mathbb{F}}$, we wish to recover K, τ, and α.

2.1 Notation

Before going any further, let us collect the notation used throughout this paper. Absolute constants are denoted by $C_1, C_2, \cdots$. In addition, C denotes a constant that might change in each appearance. We will occasionally use the convention that $[a:b] = \{a, a+1, \cdots, b\}$ for integers $a \le b$.

The standard asymptotic notation is freely used in this work and is reviewed next for the reader's convenience.

- For functions $a, b : \mathbb{C} \to \mathbb{C}$, $a(\theta) = O(b(\theta))$ asymptotically as $\theta \to \infty$ if there exists positive constants C_1 and C_2 such that

$$|a(\theta)| \le C_1 \cdot |b(\theta)|, \qquad |\theta| > C_2.$$

- We use the conventions [18] that

$$a(\theta) = \Omega(b(\theta)) \Longleftrightarrow b(\theta) = O(a(\theta)),$$

$$a(\theta) = \Theta(b(\theta)) \Longleftrightarrow a(\theta) = O(b(\theta)) \text{ and } a(\theta) = \Omega(b(\theta)),$$

  asymptotically as $\theta \to \infty$.

- Lastly, $a(\theta) = o(b(\theta))$ asymptotically as $\theta \to \infty$ if, for every $\epsilon > 0$, there exists $\delta = \delta(\epsilon)$ such that

$$|a(\theta)| \le \epsilon \cdot |b(\theta)|, \qquad |\theta| > \delta(\epsilon).$$

  In particular, as long as $\lim_{\theta\to\infty} b(\theta) \ne 0$, we have that

$$a(\theta) = o(b(\theta)) \Longleftrightarrow \lim_{\theta\to\infty} \frac{a(\theta)}{b(\theta)} = 0.$$

Recall that $\mathbb{I} = [0, 1)$. The natural norms on $L_2(\mathbb{I})$, $L_1(\mathbb{I})$, and $L_\infty(\mathbb{I})$ are denoted by the shorthands $\|\cdot\|_{L_2}$, $\|\cdot\|_{L_1}$, and $\|\cdot\|_{L_\infty}$, respectively. For the wraparound metric $d(\cdot,\cdot)$ defined in (1), the Hausdorff distance between sets $\mathbb{A}$ and $\mathbb{B}$ (both subset of $\mathbb{I}$) is defined as

$$d(\mathbb{A}, \mathbb{B}) := \max\left\{ \sup_{a\in\mathbb{A}} \inf_{b\in\mathbb{B}} d(a,b), \sup_{b\in\mathbb{B}} \inf_{a\in\mathbb{A}} d(a,b) \right\}. \tag{8}$$

Effectively, $d(\mathbb{A}, \mathbb{B})$ controls the distance from any point on $\mathbb{A}$ to $\mathbb{B}$ and vice verse. With some abuse of notation, we define the Hausdoff distance of two vectors in the natural way (as the distance between the finite sets formed by their entries).

Throughout, $\circledast$ stands for circular convolution, which corresponds to point-wise multiplication in the Fourier series domain. Lastly, to unburden the notation, we occasionally suppress the dependence on different quantities if there is no ambiguity.

3 Phase I: Initialization

Rather than recovering the amplitudes α, we focus on estimating the positions of impulses τ. Indeed, given an estimate of the positions, an estimate for the amplitudes readily follows from a simple least-squares calculation.

In this section, we present a simple iterative algorithm that, given the noisy Fourier coefficients of $x_{\tau,\alpha}$ on the interval $\mathbb{F} = [-f_C : f_C]$, approximately recovers the position vector τ under a certain separability condition (that we will specify shortly).

This algorithm requires a band-limited *kernel*. More specifically, let $N := 2f_C + 1 = |\mathbb{F}|$ for short. Then, for $\sigma \in (0, \frac{1}{2})$, we assume that the kernel $g_{\sigma,N}(\cdot) = g(\cdot;\sigma, N) : \mathbb{I} \to \mathbb{R}$ is band-limited to $\mathbb{F}$ and decays sharply away from the origin so that $|g_{\sigma,N}(t)|$ is small when $t \in [\sigma, 1-\sigma]$. The next statement formally lists the requirements on the kernel.

Criterion 1. *For integer f_C, set $N = 2f_C + 1$ for short. For $c = c(N) > 0$, let $\sigma = \frac{c}{N} < \frac{1}{2}$. The kernel $g_{\sigma,N}(\cdot) : \mathbb{I} \to \mathbb{R}$ satisfies the following requirements.*

- *First, $g_{\sigma,N}(\cdot)$ has unit-energy, $\|g_{\sigma,N}(\cdot)\|_{L_2} = 1$, and is band-limited to $\mathbb{F} = [-f_C : f_C]$ (so that the Fourier coefficients $\{\widehat{g}_{\sigma,N}[l]\}$ vanish when $l \notin \mathbb{F}$).*
- *Second, $g_{\sigma,N}(\cdot)$ is symmetric about $\frac{1}{2}$ (so that $g_{\sigma,N}(t) = g_{\sigma,N}(1-t)$ for every $t \in \mathbb{I}$).*
- *Lastly, as $c, N \to \infty$ with $c = c(N) = O(\log N)$, the decay of $g_{\sigma,N}(\cdot)$ away from the origin is asymptotically quantified as*

$$|g_{\sigma,N}(t)| = \frac{O(e^{-C_3 c})}{\sqrt{N}\sin(\pi t)}, \qquad \sigma \le t \le \frac{1}{2},$$

$$g_{\sigma,N}(0) = \Omega(1/\sqrt{\sigma}) = \Omega\left(\sqrt{\frac{N}{c}}\right),$$

for some constant $C_3 > 0$.

As mentioned earlier, the success of the Phase I algorithm, summarized in Figure 2, hinges on Criterion 1. Throughout this section, we assume the existence of a kernel $g_{\sigma_1,N}(\cdot)$ for which Criterion 1 holds with $c = c_1 = c_1(N)$ and $\sigma_1 = \frac{c_1}{N}$.

In a nutshell, Algorithm I iteratively finds the largest peaks of $z^1_{\sigma_1}(\cdot) = [g_{\sigma_{1,N}} \circledast y](\cdot)$, and in order to avoid (falsely) detecting the nearby points in the next iteration, erases the neighborhood (of radius $2\sigma_1$) of each peak. In fact, Algorithm I may loosely be considered as the extension of *orthogonal matching pursuit* to a continuous domain [20].

As we describe next, under Criterion 1, Algorithm I returns a reliable estimate of K and τ as long as the impulse locations $\{\tau[i]\}_i$ are well-separated and the dynamic range of $x_{\tau,\alpha}$ is not too large. To be concrete, the *separation* of τ is defined as follows [3]:

$$\mathrm{sep}(\tau) := \min_{\substack{i,j \in [1:K] \\ i \neq j}} d\left(\tau[i], \tau[j]\right). \tag{9}$$

In addition, we define the dynamic range of $x_{\tau,\alpha}$ as follows:

$$\mathrm{dyn}(x_{\tau,\alpha}) := \frac{\max_{i\in[1:K]} |\alpha[i]|}{\min_{i\in[1:K]} |\alpha[i]|}. \tag{10}$$

The performance guarantee for Algorithm 1 is summarized below and proved in Section 6.1.

Proposition 2. [Performance of Algorithm I] *Fix a measure $x_{\tau,\alpha}$ with number of impulses K, vector of positions $\tau \in \mathbb{I}^K$ with distinct entries, and vector of amplitudes $\alpha \in \mathbb{R}^K$ defined as in (2). With the cut-off frequency $f_C \in \mathbb{N}$ and $\mathbb{F} = [-f_C : f_C]$, let $y(\cdot)$ be the (possibly noisy) measurement signal band-limited to $\mathbb{F}$. The nonzero Fourier coefficients of $y(\cdot)$ are*

$$\widehat{y}[l] = \widehat{x}_{\tau,\alpha}[l] + \widehat{n}[l], \qquad l \in \mathbb{F},$$

Algorithm I (initialization)

Input:

- A cut-off frequency $f_C \in \mathbb{N}$ and a measurement signal $y(\cdot)$ that is band-limited to $\mathbb{F} = [-f_C : f_C]$ (see (7)).
- With $N = 2f_C + 1$ and $0 < \sigma_1 < \frac{1}{2}$, a kernel $g_{\sigma_1,N}(\cdot)$ (see Criterion 1).
- A threshold $\eta > 0$.

Output:

- An estimate of K and τ, denoted here by $\widetilde{K}$ and $\tau^0 \in \mathbb{I}^{\widetilde{K}}$, respectively.

1. Compute $z^1_{\sigma_1}(\cdot) := (g_{\sigma_1,N} \circledast y)(\cdot)$. Here, $\circledast$ stands for circular convolution.
2. Set $j = 1$. As long as $\|z^j_{\sigma_1}(\cdot)\|_{L_\infty} > \eta$, repeat the following (where $d(\cdot,\cdot)$ is the wraparound metric defined in (1)):
   (a) $\tau^0[j] = \arg\max_{t\in\mathbb{I}} \left|z^j_{\sigma_1}(t)\right|$.
   (b) $z^{j+1}_{\sigma_1}(t) = \begin{cases} z^j_{\sigma_1}(t) & d\left(t, \tau^0[j]\right) > 2\sigma_1, \\ 0 & d\left(t, \tau^0[j]\right) \le 2\sigma_1. \end{cases}$
   (c) $j \leftarrow j + 1$.
3. Set $\widetilde{K} = j - 1$ to be the estimate of number of impulses K. Also, return $\tau^0 \in \mathbb{I}^{\widetilde{K}}$ as the estimate of their locations τ.

Figure 2: Algorithm I (initialization)

where $\widehat{x}_{\tau,\alpha}$ and $\widehat{n}$ are the Fourier series of $x_{\tau,\alpha}$ and the noise $n(\cdot)$, respectively (see (7)).

For $N = 2f_C + 1$ and $c_1 = c_1(N) > 0$, set $\sigma_1 = \frac{c_1}{N} < \frac{1}{2}$. In what follows, $c_1, N \to \infty$, and $c_1 = \Theta(\log N)$ with a sufficiently large lower bound.[7] *Suppose that the kernel $g_{\sigma_1,N}(\cdot)$ satisfies Criterion 1, and that the threshold η in Algorithm I is specified as*

$$\eta = 2\|n(\cdot)\|_{L_\infty}.$$

Then, the output of Algorithm I assymptotically (i.e., for large enough N)[8] *satisfies the following:*

- *$\widetilde{K} = K$, i.e., Algorithm I correctly estimates the number of impulses, and*
- *$d(\tau^0, \tau) \le \sigma_1$, i.e., the Hausdorff distance between the vector of true positions τ and the estimates returned by Algorithm I is small.*

[7] That is, for large enough factors $\alpha \le \beta$ specified in the proof and for sufficiently large N, we assume that $\alpha \log N \le c_1 \le \beta \log N$.

[8] In particular, N must be large enough so that $\text{sep}(\tau) \ge 4\sigma_1$ and $\text{dyn}(x_{\tau,\alpha}) = O(\text{NSR}/\sqrt{\sigma_1})$, where NSR stands for the noise-to-signal ratio (as specified in (45)).

Remark 3. Algorithm I returns an initial estimate of $\tau \in \mathbb{I}^K$, namely $\tau^0 \in \mathbb{I}^K$. In the second part of this work, we refine this initial estimate by solving a local optimization program. In particular, asymptotically, we will be able to recover τ exactly from noise-free low-frequency measurements.

4 Phase II: Local Optimization

This section presents a method to refine the estimate of τ produced by Algorithm I (namely, $\tau^0 \in \mathbb{I}^{\widetilde{K}}$). In particular, asymptotically and in the absence of noise, we will be able to exactly recover τ.

With $c_2 = c_2(N) > 0$ to be specified later, set $\sigma_2 = \frac{c_2}{N}$. For cut-off frequency f_C, recall that $\mathbb{F} = [-f_C : f_C]$ and that $N = |\mathbb{F}| = 2f_C + 1$. Throughout this section, we consider the unit-energy kernel $g_{\sigma_2,N}(\cdot)$ which is band-limited to $\mathbb{F}$ by design. We first filter the low-frequency measurement signal $y(\cdot)$ with the kernel $g_{\sigma_2,N}(\cdot)$. More precisely, for $t \in \mathbb{I}$, we set

$$\begin{aligned} z_{\sigma_2}(t) &:= (g_{\sigma_2,N} \circledast y)(t) \\ &= (g_{\sigma_2,N} \circledast x_{\tau,\alpha})(t) + (g_{\sigma_2,N} \circledast n)(t) \qquad \text{(see (7) and the text below)} \\ &= \sum_{i=1}^{K} \alpha[i] \cdot g_{\sigma_2,N}(t \ominus \tau[i]) + (g_{\sigma_2,N} \circledast n)(t) \qquad \text{(see (2))} \\ &=: \sum_{i=1}^{K} \alpha[i] \cdot g_{\sigma_2,N}(t \ominus \tau[i]) + n_{\sigma_2}(t), \qquad (n_{\sigma_2}(\cdot) \text{ is the filtered noise}) \end{aligned} \tag{11}$$

where the second line above uses the assumption that $g_{\sigma_2,N}(\cdot)$ is also band-limited to $\mathbb{F}$. Let $\widehat{g}_{\sigma_2,N}$ and $\widehat{n}_{\sigma_2}$ be the corresponding Fourier series. Note that $\widehat{g}_{\sigma_2,N}$ is supported only on the interval $\mathbb{F}$ and so is $\widehat{n}_{\sigma_2}$. In light of (11), the Fourier coefficients of $z_{\sigma_2}(\cdot)$ can then be written as

$$\begin{aligned} \widehat{z}_{\sigma_2}[l] &= \widehat{g}_{\sigma_2,N}[l] \cdot \widehat{y}[l] \\ &= \widehat{g}_{\sigma_2,N}[l] \sum_{i=1}^{K} \alpha[i] e^{-\mathrm{i}\,2\pi l \tau[i]} + \widehat{n}_{\sigma_2}[l], \qquad l \in \mathbb{F}, \end{aligned} \tag{12}$$

where the second line follows from a direct calculation. For a more compact representation, we abuse the notation by letting $\widehat{z}_{\sigma_2}, \widehat{n}_{\sigma_2} \in \mathbb{C}^N$ also denote the vectors formed by the Fourier series coefficients of $z_{\sigma_2}(\cdot)$ and $n_{\sigma_2}(\cdot)$ on $\mathbb{F}$, respectively. Then, the vector form of (12) is simply

$$\widehat{z}_{\sigma_2} = G_\tau \cdot \alpha + \widehat{n}_{\sigma_2}, \tag{13}$$

where $G_\tau \in \mathbb{C}^{N \times K}$ is constructed out of the (modulated) Fourier coefficients of the kernel:

$$G_\tau[l,i] = \widehat{g}_{\sigma_2,N}[l] \cdot e^{-\mathrm{i}\,2\pi l \cdot \tau[i]}, \qquad l \in \mathbb{F},\ i \in [1:K]. \tag{14}$$

Alternatively, $G_\tau \alpha \in \mathbb{C}^N$ is the vector formed by the Fourier series coefficients of the filtered measure $(g_{\sigma_2,N} \circledast x_{\tau,\alpha})(\cdot)$ on $\mathbb{F}$.

In this section, we will rely on $g_{\sigma_2,N}(\cdot)$ satisfying a number of properties. The first criterion, among other things, specifies how small the correlation between the kernel and its shifted copy should be (when separated properly). The second criterion concerns the behavior of kernel near the origin (and that it must be "flat" in a very small interval near the origin).

Criterion 4. *For integer f_C, set $N = 2f_C + 1$ for short. For $c = c(N) > 0$, let $\sigma = \frac{c}{N} < \frac{1}{2}$. The kernel $g_{\sigma,N}(\cdot)$ has unit energy $\|g_{\sigma,N}(\cdot)\|_{L_2} = 1$, is band-limited to $\mathbb{F}$, and symmetric about $\frac{1}{2}$ (as in Criterion 1). Moreover, $g_{\sigma,N}(\cdot)$ satisfies the following.*

For $\rho_1, \rho_2 \in \mathbb{I}$ with $d(\rho_1, \rho_2) \geq 2\sigma$, it holds asymptotically that

$$|\langle g_{\sigma,N}(t \ominus \rho_1), g_{\sigma,N}(t \ominus \rho_2)\rangle| = \frac{O(e^{-C_3 c})}{N \cdot \sin(\pi \cdot d(\rho_1, \rho_2))}, \tag{15}$$

$$\left|\left\langle g_{\sigma,N}(t \ominus \rho_1), g'_{\sigma,N}(t \ominus \rho_2)\right\rangle\right| = \frac{O(e^{-C_3 c})}{\sin(\pi \cdot d(\rho_1, \rho_2))}, \tag{16}$$

$$\left|\left\langle g'_{\sigma,N}(t \ominus \rho_1), g'_{\sigma,N}(t \ominus \rho_2)\right\rangle\right| = \frac{N \cdot O(e^{-C_3 c})}{\sin(\pi \cdot d(\rho_1, \rho_2))}, \tag{17}$$

when $c, N \to \infty$ and $c = O(\log N)$. Here, $d(\rho_1, \rho_2)$ is the wraparound distance between ρ_1 and ρ_2 (see (1)), and $g'_{\sigma,N}(\cdot)$ denotes the derivative of $g_{\sigma,N}(\cdot)$ with respect to its argument.

Criterion 5. *For integer f_C, set $N = 2f_C + 1$ for short. For $c = c(N) > 0$, let $\sigma = \frac{c}{N} < \frac{1}{2}$. There exists $h(\sigma, N) \leq \sigma$ (depending only on σ and N), for which the kernel $g_{\sigma,N}(\cdot)$ satisfies the following.*

For $\rho_1, \rho_2 \in \mathbb{I}$, suppose that $d(\rho_1, \rho_2) \leq h(\sigma, N) \leq \sigma$. Then, it holds asymptotically that $\|g'_{\sigma,N}(\cdot)\|^2_{L_2} = \Omega(N^2)$, and furthermore

$$\langle g_{\sigma,N}(t \ominus \rho_1), g_{\sigma,N}(t \ominus \rho_2)\rangle = 1 - O(1) \cdot d^2(\rho_1, \rho_2), \tag{18}$$

$$\left|\left\langle g_{\sigma,N}(t \ominus \rho_1), g'_{\sigma,N}(t \ominus \rho_2)\right\rangle\right| = \Omega\left(N^2\right) \cdot d(\rho_1, \rho_2), \tag{19}$$

$$\mathit{sign}\left(\left\langle g_{\sigma,N}(t \ominus \rho_1), g'_{\sigma,N}(t \ominus \rho_2)\right\rangle\right) = \mathit{sign}\left(\rho_1 \ominus \rho_2 - \frac{1}{2}\right), \tag{20}$$

as $c, N \to \infty$ with $c = O(\log N)$. Above, sign$(\cdot)$ returns the sign, of course.

Throughout this section, we assume that the kernel $g_{\sigma_2,N}(\cdot)$ satisfies both Criteria 4 and 5 with $\sigma = \sigma_2 = \frac{c_2}{N}$. We will soon specify $c_2 = c_2(N)$ in relation to c_1 (from Phase I).

Define $G_\rho \in \mathbb{C}^{N \times \widetilde{K}}$ similar to (14) (but with $\widetilde{K}$ instead of K), and consider the objective function

$$f(\rho, \beta) := \|G_\rho \beta - \widehat{z}_{\sigma_2}\|_2^2, \qquad \rho \in \mathbb{I}^{\widetilde{K}},\ \beta \in \mathbb{R}^{\widetilde{K}}, \tag{21}$$

with ρ and β being vectors of positions and amplitudes, respectively. For a fixed $\rho \in \mathbb{I}^{\widetilde{K}}$, minimizing $f(\rho, \cdot)$ is a simple least-squares problem:

$$\begin{aligned}
\min_{\rho \in \mathbb{I}^{\widetilde{K}}} \min_{\beta \in \mathbb{R}^{\widetilde{K}}} f(\rho, \beta) &= \min_{\rho \in \mathbb{I}^{\widetilde{K}}} \min_{\beta \in \mathbb{R}^{\widetilde{K}}} \|G_\rho \beta - \widehat{z}_{\sigma_2}\|_2^2 \\
&= \min_{\rho \in \mathbb{I}^{\widetilde{K}}} f(\rho, \beta_\rho) \qquad \left(\beta_\rho := G_\rho^\dagger \cdot \widehat{z}_{\sigma_2}\right) \\
&= \min_{\rho \in \mathbb{I}^{\widetilde{K}}} \|(I_N - \mathcal{P}_\rho)\, \widehat{z}_{\sigma_2}\|_2^2 \qquad \left(\mathcal{P}_\rho = G_\rho G_\rho^\dagger \in \mathbb{C}^{N \times N}\right) \\
&=: \min_{\rho \in \mathbb{I}^{\widetilde{K}}} F(\rho).
\end{aligned} \tag{22}$$

Above, $G_\rho^\dagger$ is the Moore-Penrose pseudo-inverse of $G_\rho \in \mathbb{C}^{N \times \widetilde{K}}$. Also, $\mathcal{P}_\rho = G_\rho G_\rho^\dagger$ is the orthogonal projection onto $\mathrm{span}(G_\rho)$.

Suppose that Proposition 2 is in force so that, in particular, $\widetilde{K} = K$. Now, (13) suggests that minimizing $F(\cdot)$ in (22) might reliably estimate the true vector of positions $\tau \in \mathbb{I}^K$. In fact, in the absence of noise, τ is indeed a solution to Program (22) (with $\widetilde{K} = K$).[9]

However, even in the absence of noise, the super-resolution problem (Program (22)) might have multiple local minima in which an optimization algorithm might get trapped. The key insight that resolves this issue is that, under Proposition 2, the outcome of Algorithm I (namely, $\tau^0 \in \mathbb{I}^{\widetilde{K}}$ with $\widetilde{K} = K$) is close enough to τ so that a local optimization algorithm (initialized at τ^0) converges to τ (or its small vicinity).

To formalize matters, we cast the local optimization step as follows. Under Proposition 2, recall that $d(\tau^0, \tau) \leq \sigma_1$. To incorporate this prior knowledge, we add this constraint to Program (22) to obtain the box-constrained program

$$\min_{\rho \in \mathbb{B}(\tau^0, \sigma_1)} F(\rho), \tag{23}$$

where

$$\mathbb{B}(\tau^0, \sigma_1) := \left\{ \rho \in \mathbb{I}^{\widetilde{K}} \, : \, d(\rho, \tau^0) \leq \sigma_1 \right\} \subset \mathbb{I}^{\widetilde{K}}, \tag{24}$$

is a ball of radius σ_1 centered at the initial estimate τ^0 from Algorithm I.[10] Given τ^0, one might use any constrained optimization algorithm to solve Program (23).

Before discussing two such algorithms, let us shed light on the geometry of the ball $\mathbb{B}(\tau^0, \sigma_1)$, when the entries of τ^0 are distinct,[11] whereby

$$\mathbb{B}(\tau^0, \sigma_1) = \left\{ \rho \in \mathbb{I}^{\widetilde{K}} \, : \, d(\rho[i], \tau^0[i]) \leq \sigma_1, \quad i \in [1 : \widetilde{K}] \right\}.$$

when N is sufficiently large. If the entries of τ^0 are distinct *and* away from the origin, then we have the simpler expression

$$\mathbb{B}(\tau^0, \sigma_1) = \left[\tau^0[1] - \sigma_1, \tau^0[1] + \sigma_1 \right] \times \cdots \times \left[\tau^0[\widetilde{K}] - \sigma_1, \tau^0[\widetilde{K}] + \sigma_1 \right]. \tag{25}$$

Furthermore, the set of *active coordinates* for $\rho \in \mathbb{B}(\tau^0, \sigma_1)$ consists of the coordinates on the boundary of $\mathbb{B}(\tau^0, \sigma_1)$, that is

$$\mathbb{A}(\rho) := \left\{ i \, : \, d\left(\rho[i], \tau^0[i]\right) = \sigma_1 \right\} \subseteq [1 : \widetilde{K}], \qquad \rho \in \mathbb{B}(\tau^0, \sigma_1). \tag{26}$$

When (25) holds, for instance, $i \in \mathbb{A}(\rho)$ if simply $\rho[i] = \tau^0[i] \pm \sigma_1$. Naturally, $\mathbb{A}^C(\rho)$ (namely, the complement of $\mathbb{A}(\rho)$) consists of *inactive coordinates* of ρ.

We now turn to the details of solving Program (23). The *gradient projection algorithm* is an obvious candidate for a first-order method here.[12] At iteration $j \geq 1$, one sets

$$\tau^j = \mathcal{P}_{\mathbb{B}(\tau^0, \sigma_1)} \left(\tau^{j-1} - \delta^j \cdot \frac{\partial F}{\partial \rho}(\tau^{j-1}) \right) \in \mathbb{I}^{\widetilde{K}}, \tag{27}$$

with step size $\delta^j > 0$ at the jth iteration. Above, $\mathcal{P}_{\mathbb{B}(\tau^0, \sigma_1)}(\cdot)$ (namely, the projection operator onto the ball $\mathbb{B}(\tau^0, \sigma_1)$) ensures that τ^j remains a feasible point of Program (23) at the jth iteration. We in fact find an explicit expression for the gradient of $F(\cdot)$ in the supporting document [11]:

$$\frac{\partial F}{\partial \rho}(\rho) = -2 \cdot \text{diag}(\beta_\rho) \cdot G_\rho^* L \left(I_N - \mathcal{P}_\rho \right) \widehat{z}_{\sigma_2} \in \mathbb{R}^{\widetilde{K}}, \qquad \rho \in \mathbb{I}^{\widetilde{K}}. \tag{28}$$

[9] In general, given an estimate $\widetilde{\tau} \in \mathbb{I}^{\widetilde{K}}$ of $\tau \in \mathbb{I}^K$, an estimate of amplitudes $\alpha \in \mathbb{R}^K$ is simply $\beta_{\widetilde{\tau}} = G_{\widetilde{\tau}}^\dagger \cdot \widehat{z}_{\sigma_2} \in \mathbb{R}^{\widetilde{K}}$.

[10] Note that we used $\widetilde{K}$ to define the ball (instead of K), so as to develop Phase II independent of Proposition 2. The theoretical guarantees for Phase II, however, do indeed depend on the success of Phase I. Specifically, when it comes to the theory of Phase II, we will assume that Proposition 2 is in force: $\tau^0 \in \mathbb{I}^{\widetilde{K}}$ with $\widetilde{K} = K$, and $d(\tau^0, \tau) \leq \sigma_1$.

[11] For example, under Proposition 2, the entries of τ^0 are distinct asymptotically, i.e., for sufficiently large N.

[12] Alternatively, one may use the *conditional gradient method* instead of the gradient projection algorithm [16].

Here, $L \in \mathbb{C}^{N\times N}$ is a diagonal matrix with $L[l,l] = \mathrm{i}\, 2\pi l$, $l \in \mathbb{F}$. Also, $\beta_\rho = G_\rho^\dagger \cdot \widehat{z}_{\sigma_2}$ and $\mathcal{P}_\rho = G_\rho G_\rho^\dagger$. Moreover, $\mathrm{diag}(\beta_\rho)$ is the diagonal matrix formed by the vector β_ρ. Without a formal proof we remark that the gradient projection algorithm converges to τ (or its neighborhood when there is noise).

A valuable fact here is that once the gradient projection algorithm identifies an active coordinate, that coordinate remains unchanged in future iterations. More specifically, if $d(\tau^j[i], \tau^0[i]) = \sigma_1$, then $d(\tau^{j'}[i], \tau^0[i]) = \sigma_1$ for all future iterations $j' \geq j$ [16].[13]

From a practical standpoint, however, deploying a first-order method (such as the gradient projection algorithm above) is somewhat unwise since the initial estimate τ^0 is generally too close to τ and, as a result, $\frac{\partial F}{\partial \rho}(\tau^0) \approx 0$. This in turn results in a slow—linear to be precise—convergence rate.

Actually, the local nature of this problem encourages second-order methods as a viable alternative here. To proceed, let

$$\frac{\partial^2 F}{\partial \rho^2}(\rho) = \left[\frac{\partial^2 F}{\partial \rho[i] \partial \rho[j]}(\rho)\right]_{i,j} \in \mathbb{R}^{\widetilde{K}\times\widetilde{K}}$$

denote the Hessian of $F(\cdot)$ at $\rho \in \mathbb{I}^{\widetilde{K}}$, and define the *reduced Hessian* at ρ to be

$$\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni \mathcal{R}\left(\frac{\partial^2 F}{\partial \rho^2}(\rho)\right) = \begin{cases} \delta_{i,j} & i \in \mathbb{A}(\rho) \,\mathrm{or}\, j \in \mathbb{A}(\rho), \\ \frac{\partial^2 F}{\partial \rho[i]\partial\rho[j]}(\rho) & \text{otherwise.} \end{cases} \tag{29}$$

Here, $\delta_{i,j}$ is the Kronecker delta function, $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$. Because $F(\cdot)$ is a smooth function, its reduced Hessian is positive semi-definite near a solution $\widetilde{\tau}$ of Program (23) [16], i.e.,

$$\mathcal{R}\left(\frac{\partial^2 F}{\partial \rho^2}(\rho)\right) \succcurlyeq 0, \qquad \text{when } d\left(\rho, \widetilde{\tau}\right) \text{ is small.}$$

Therefore, hypothetically, if $\mathbb{A}(\widetilde{\tau})$ (namely, the active coordinates of $\widetilde{\tau}$) were known and $d(\tau^0, \widetilde{\tau})$ was small, we could have calculated the rest of coordinates of $\widetilde{\tau}$ by applying the basic unconstrained Newton's method (using the reduced Hessian in (29) and assuming its invertibility).

Of course, we will not know the active constraints until the problem is solved. Instead, we must use the *projected Newton's method*. In words, at each iteration, the projected Newton's method carefully underestimates the active coordinates. Then, the (estimated) inactive coordinates are updated using an (unconstrained) Newton's step, and the (estimated) active coordinates are in turn updated using the gradient projection step. Loosely speaking, the fact that active coordinates remain unchanged under the gradient projection algorithm is the key to the success of projected Newton's method.

To formally write down the iterations of the projected Newton's method [16, Algorithm 5.5.2], we record a couple more definitions. For $\varepsilon > 0$, the ϵ-active coordinates of ρ are collected in the set

$$\mathbb{A}_\epsilon(\rho) := \left\{i \,:\, \sigma_1 - \epsilon \leq d\left(\rho[i], \tau^0[i]\right) \leq \sigma_1\right\} \subseteq [1:\widetilde{K}], \qquad \rho \in \mathbb{B}(\tau^0, \sigma_1). \tag{30}$$

In particular, $\mathbb{A}_0(\rho) = \mathbb{A}(\rho)$ (see 26). The ϵ-reduced Hessian is defined similar to (29) as

$$\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni \mathcal{R}_\epsilon\left(\frac{\partial^2 F}{\partial \rho^2}(\rho)\right) = \begin{cases} \delta_{i,j} & i \in \mathbb{A}_\epsilon(\rho) \text{ or } j \in \mathbb{A}_\epsilon(\rho), \\ \frac{\partial^2 F}{\partial \rho[i]\partial\rho[j]}(\rho) & \text{otherwise.} \end{cases} \tag{31}$$

[13] A similar phenomenon is true of any convex feasible set (and not just box constraints).

Also, let us give an explicit (if not elegant) expression for the Hessian (which is verified in the accompanying document [11]):

$$\begin{aligned}
\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni \frac{\partial^2 F}{\partial \rho^2}(\rho) = &-2\cdot \operatorname{diag}(\beta_\rho)\cdot G_\rho^* L^2 G_\rho \cdot \operatorname{diag}(\beta_\rho) \\
&- 2\cdot \operatorname{diag}(\beta_\rho)\cdot \operatorname{diag}\left(G_\rho^* L^2 (I_N - \mathcal{P}_\rho)\, \widehat{z}_{\sigma_2}\right) \\
&- 2\left[\operatorname{diag}(\beta_\rho)\, G_\rho^* L G_\rho - \operatorname{diag}\left(G_\rho^* L (I_N - \mathcal{P}_\rho)\, \widehat{z}_{\sigma_2}\right)\right] \\
&\cdot \left(G_\rho^* G_\rho\right)^{-1} \cdot \left[G_\rho^* L^* G_\rho \cdot \operatorname{diag}(\beta_\rho) - \operatorname{diag}\left(G_\rho^* L (I_N - \mathcal{P}_\rho)\, \widehat{z}_{\sigma_2}\right)\right]. \qquad (32)
\end{aligned}$$

The quantities involved (β_ρ, G_ρ, L, $\mathcal{P}_\rho$, and $\widehat{z}_{\sigma_2}$) were defined earlier. Algorithm II (in Figure 3) describes how to refine the initial estimate $\tau^0 \in \mathbb{I}^{\widetilde{K}}$ using the projected Newton's algorithm.

Algorithm II (local optimization)

Input:

- Cut-off frequency f_C and measurement signal $y(\cdot)$, band-limited to $\mathbb{F} = [-f_C : f_C]$, with the corresponding Fourier coefficients collected in $\widehat{y} \in \mathbb{C}^N$ (with $N = 2f_C + 1$). (See (5) and (7).)
- A kernel $g_{\sigma_2,N}(\cdot)$, band-limited to $\mathbb{F}$, with the corresponding Fourier coefficients collected in $\widehat{g}_{\sigma_2,N} \in \mathbb{C}^N$. (See Criteria 4 and 5.)
- From Algorithm I, an initial estimate of the vector of locations $\tau^0 \in \mathbb{I}^{\widetilde{K}}$, and $\sigma_1 \in (0, \frac{1}{2})$.
- A margin $0 < \epsilon^0 < \sigma_1$, and a termination threshold $\eta > 0$.

Output:

- An estimate $\widetilde{\tau} \in \mathbb{I}^{\widetilde{K}}$ of the true vector of locations τ.

1. Compute $\widehat{z}_{\sigma_2} = \widehat{g}_{\sigma_2,N} \odot \widehat{y}$. Here, $\odot$ stands for entry-wise (Hadamard) product.
2. Set $j = 0$ and repeat:
   (a) Compute the gradient of $F(\cdot)$ at $\tau^j \in \mathbb{I}^{\widetilde{K}}$ (see (28)).
   (b) Calculate the reduced Hessian of $F(\cdot)$ at τ^j, i.e., $\mathcal{R}_{\epsilon^j}(\frac{\partial^2 F}{\partial \rho^2}(\tau^j))$. (See (30-32).) If the reduced Hessian is not a positive definite matrix, exit with a failure message.
   (c) Calculate the descent direction $\mathbb{R}^{\widetilde{K}} \ni v^j := \left(\mathcal{R}_{\epsilon^j}\left(\frac{\partial^2 F}{\partial \rho^2}(\tau^j)\right)\right)^{-1} \cdot \frac{\partial F}{\partial \rho}(\tau^j)$.
   (d) For $\lambda \geq 0$, set $\tau^j(\lambda) = \mathcal{P}_{\mathbb{B}(\tau^0,\sigma_1)}\left(\tau^j \ominus \lambda \cdot v^j\right)$, where $\mathcal{P}_{\mathbb{B}(\tau^0,\sigma_1)}(\cdot)$ is the projection onto the ball $\mathbb{B}(\tau^0, \sigma_1)$ (see (24)).
   (e) If $\|\tau^j(1) - \tau^j\|_2 \leq \eta$, exit. Otherwise, pick $\epsilon^{j+1} = \min[\|\tau^j(1) - \tau^j\|_2, \sigma_1]$.
   (f) *Line search:* Find the least integer m such that $F\left(\tau^j(\lambda)\right) - F\left(\tau^j\right) \leq \frac{-10^{-4}}{\lambda}\left\|\tau^j(\lambda) \ominus \tau^j\right\|_2^2$, holds for $\lambda = 2^{-m}$.
   (g) Set $\tau^{j+1} = \tau^j(2^{-m})$.
   (h) $j \leftarrow j + 1$.
3. Output $\widetilde{\tau} = \tau^j \in \mathbb{I}^{\widetilde{K}}$ as the estimate of the true location vector τ.

Figure 3: Algorithm II (local optimization)

Under Criteria 4 and 5, and when Proposition 2 is in force, Algorithm II successfully refines our estimate of the true position vector τ. Convergence of the projected Newton's algorithm to τ (or its small vicinity) is guaranteed by the next result, which is proved in Section 6.2. We remark that, while not the focus of this work, similar guarantees hold for the gradient projection algorithm outlined in (27).

Theorem 6. [Performance of Algorithm II] *For integer N and $0 < c_1 < c_2$ (both functions of*

N), let $\sigma_1 = \frac{c_1}{N}$ and $\sigma_2 = \frac{c_2}{N}$. Let $\tau^0 \in \mathbb{I}^{\widetilde{K}}$ be the output of Algorithm I, and suppose that Proposition 2 is in force, so that in particular $\widetilde{K} = K$. Suppose also that the kernel $g_{\sigma_2,N}(\cdot)$ satisfies Criteria 4 and 5 (with $\sigma = \sigma_2$). Lastly, assume that $2\sigma_1 \le h(\sigma_2, N) \le \sigma_2$ (see Criterion 5).

Then, as long as

$$\frac{\|n(\cdot)\|_{L_2}}{\|\alpha\|_\infty} = \frac{O(1)}{dyn\,(x_{\tau,\alpha})^2} \le 1,$$

with a small enough constant, any limit point of Algorithm II is a stationary point $\widetilde{\tau} \in \mathbb{I}^{\widetilde{K}}$ with $\widetilde{K} = K$, and

$$d(\widetilde{\tau}, \tau) \le \min\left(O(1) \cdot dyn(x_{\tau,\alpha})^2 \cdot \frac{\|n(\cdot)\|_{L_2}}{\|\alpha\|_2}, 2\sigma_1\right), \tag{33}$$

asymptotically as $c_1, c_2, N \to \infty$ and $c_1, c_2 = \Theta(\log N)$ (with a large enough lower bound). Above, the metric and the dynamic range $dyn(x_{\tau,\alpha})$ were defined in (8) and (10), respectively, and $\|n(\cdot)\|_{L_2}$ is the energy of the additive noise (see (4)).

A few remarks are in order.

Remark 7. **[Noise-free]** From (33), we observe that Phase II refines the output of Phase I when the dynamic range and noise level are both moderate. In particular, in the absence of noise, Phase II exactly identifies the correct support: $\widetilde{\tau} = \tau$.

Remark 8. **[Separation]** For the two-phase algorithm to succeed (i.e., for Proposition 2 and Theorem 6 to hold), the spike locations should be well-separated. In particular, for sufficiently large f_C, one needs

$$\mathrm{sep}(\tau) \ge 4\sigma_1 = \Omega(1) \cdot \frac{\log f_C}{f_C} \tag{34}$$

(as indicated in Proposition 2).

In contrast, super-resolution via convex relaxation requires a separation of $\Omega(1/f_C)$ [3]. It is not clear whether the extra logarithmic factor in (34) is an artifact of the proofs of Proposition 2 or Theorem 6. We also recently learned about similar rates (obtained with different techniques) in the context of edge detection from limited Fourier measurements [5]. It appears that further work is needed to find possible connections and to determine whether the required separation in (34) is optimal.

Remark 9. **[Computational complexity]** As mentioned earlier, the two-phase algorithm for super-resolution is very fast, in part because fast and convenient means for generating the kernels (namely, DPSWFs, which we recommend) exist, and partly because the search space in Phase II is K-dimensional where K (the number of impulses) is often small (see Program (23)). Also confer Section 5.

5 Prior Art

By leveraging the *sparsity* of the signal model in (2), Candès et al. [3] proposed a super-resolution algorithm that involves solving a convex program—a (typically expensive) SDP to be precise. In the absence of noise, this SDP precisely recovers the sparse measure $x_{\tau,\alpha}$. More generally, the energy of the smoothed error signal scales with the noise level [2]. Later, these results were translated into bounds on the distance between the estimated and true impulse positions [13]. We remark that [3] was followed by several good papers, including [26, 8, 1, 14, 7, 19], that either proposed new super-resolution algorithms or improved the computational complexity and performance of existing methods.

But perhaps [12] is more relevant to the present work. There, Fannjiang et al. modified the orthogonal matching pursuit algorithm to handle the highly coherent over-sampled DFT matrix. To improve the robustness of the algorithm, a local optimization step is skillfully implemented in each step of their algorithm. This step refines one impulse position $\tau[i]$ at a time while keeping the rest of τ fixed. The present work differs from [12] in its use of prolate functions, and in the depth of its theoretical guarantees. In particular, [12] does not seem to offer an analogue of Theorem 6.

For the sake of demonstration, we compared our algorithm with those in [3, 12]. Each $x_{\tau,\alpha}$ was generated with number of impulses $K = 14$,[14] uniformly random positions $\tau \in \mathbb{I}^K$, and amplitudes $\alpha \in \mathbb{R}^K$ drawn independently from zero-mean Gaussian distribution with variance $(2f_C + 1)^{-1}$. Additionally, we made sure that the impulse positions were well-separated: $\text{sep}(\tau) \geq 2/f_C$ for every $x_{\tau,\alpha}$. The cut-off frequency was set to $f_C = 50$, and we set $\sigma_1 = \frac{3/2}{2f_C+1}$ and $\sigma_2 = 3\sigma_1/2$ in our algorithm. Additive low-pass Gaussian noise with energy $(2f_C + 1)\nu^2$ was then added to the observations. Figure 4 compares the (Hausdorff) distance of the estimated and true impulse positions for various values of ν, and the run-times of the algorithms.

In about 9% of the noise-free trials, the two-phase algorithm failed to exactly recover the impulse positions (but the error was still very small). In these trials, the initial estimate (output of Algorithm I) was not sufficiently close to the true impulse positions and, as a result, the local optimization phase (Algorithm II) converged to a local (as opposed to global) minimum. Recall that, according to Remark 8, the two-phase algorithm requires a separation of nearly $\log(f_C)/f_C$ to succeed (in contrast to the separation of $2/f_C$ is this experiment).

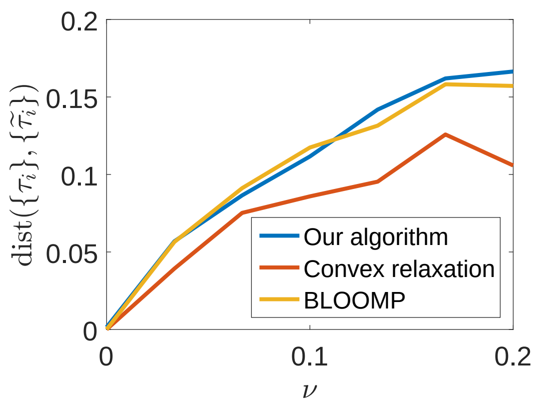

Figure 4: Comparing our algorithm to the super-resolution algorithms in [3, 12]: the horizontal axis reflects the noise level and the vertical axis displays the error, namely the distance between the estimated and true impulse positions. The average run-time for our algorithm, Fannjiang's, and Candès' were 0.4, 4, and 8.3 seconds, respectively on a laptop computer. (We made no attempts to optimize our code.)

The super-resolution problem in this paper and the problem of line spectral estimation are closely related (once the time and frequency domains are exchanged) [27, 25]. We particularly recognize Thomson's multitaper algorithm for spectral estimation [27] due to its use of prolate functions and its popularity. In Thomson's algorithm, to lower the estimation bias, data is passed through multiple *tapers*. The spectra of different channels are then averaged (often with weights)

[14]For a fair comparison, we assumed that K is known in advance so as to match the setup of [12].

to estimate the spectrum of the underlying random process (that generated the data). Because of their finite support, orthogonality, and negligible *spectral leakage*, the Fourier series of the DPSWFs (also known as DPSSs) constitute an ideal choice for the tapers. Beyond these commonalities, our work is set apart from [27] in its particular model (combination of impulses), different operating regimes (diminishing σ here versus fixed σ in [27]), and the strong supporting theory provided here.

6 Theory

6.1 Proof of Proposition 2 (Phase I)

Asymptotically (i.e., for large enough N), it holds that

$$\operatorname{sep}(\tau) \geq 4\sigma_1 = \frac{4c_1}{N}, \tag{35}$$

because the entries of τ are assumed to be distinct. That is to say that τ is asymptotically well-separated for our purposes here, as we see shortly. For $t \in \mathbb{I}$, we next observe that

$$\begin{aligned}
z^1_{\sigma_1}(t) &= (g_{\sigma_1,N} \circledast y)(t) \qquad \text{(see Algorithm I)} \\
&= (g_{\sigma_1,N} \circledast x_{\tau,\alpha})(t) + (g_{\sigma_1,N} \circledast n)(t) \\
&= \sum_{i=1}^{K} \alpha[i] \cdot g_{\sigma_1,N}(t \ominus \tau[i]) + (g_{\sigma_1,N} \circledast n)(t) \qquad \text{(see (2))} \\
&=: \sum_{i=1}^{K} \alpha[i] \cdot g_{\sigma_1,N}(t \ominus \tau[i]) + n_{\sigma_1}(t).
\end{aligned} \tag{36}$$

The second line above holds because, by assumption, $g_{\sigma_1,N}(\cdot)$ too is band-limited to $\mathbb{F} = [-f_C : f_C]$.

Under Criterion 1, the fast decay of the kernel $g_{\sigma_1,N}(\cdot)$ guarantees that $z^1_{\sigma_1}(t)$ is small when t is away from the impulses and large otherwise. Indeed, for $t \in \mathbb{I}$, whenever

$$\min_i d(t, \tau[i]) \geq \sigma_1,$$

we argue as follows. Without loss of generality, let $\tau[1]$ be the location of the closest impulse to t, $\tau[2]$ the second closest impulse, and so on. Then the fact that $\operatorname{sep}(\tau) \geq 2\sigma_1$ (asymptotically) implies that

$$d(t, \tau[i]),\ d(t, \tau[i+1]) \geq i \cdot \sigma_1, \qquad i \in [1 : K] \text{ and odd.} \tag{37}$$

Then it follows from (36) and Criterion 1 that

$$\begin{aligned}
\left|z^1_{\sigma_1}(t)\right| &\leq \max_i |\alpha[i]| \cdot \sum_{i=1}^{K} \left|g_{\sigma_1,N}(t \ominus \tau[i])\right| + \|n_{\sigma_1}(\cdot)\|_{L_\infty} \\
&= \max_i |\alpha[i]| \cdot \frac{O(e^{-C_3 c_1})}{\sqrt{N}} \sum_{i=1}^{K} \frac{1}{\sin(\pi(t \ominus \tau[i]))} + \|n_{\sigma_1}(\cdot)\|_{L_\infty} \\
&\leq \max_i |\alpha[i]| \cdot \frac{O(e^{-C_3 c_1})}{\sqrt{N}} \sum_{i=1}^{K} \frac{1}{\sin(\pi \cdot d(t, \tau[i]))} + \|n_{\sigma_1}(\cdot)\|_{L_\infty} \qquad \text{(see (1))} \\
&\leq \max_i |\alpha[i]| \cdot \frac{O(e^{-C_3 c_1})}{\sqrt{N}} \sum_{0 < i\sigma_1 \leq \frac{1}{2}} \frac{1}{\sin(\pi i \sigma_1)} + \|n_{\sigma_1}(\cdot)\|_{L_\infty}, \qquad \text{(see (37))}
\end{aligned} \tag{38}$$

asymptotically. We can further simplify the bound above by asymptotically controlling the summation in the last line as follows:

$$
\begin{aligned}
&\sum_{0<i\sigma_1\leq\frac{1}{2}} \frac{1}{\sin\left(\pi i\sigma_1\right)} \\
&\leq \frac{1}{\sin\left(\pi\sigma_1\right)} + \sigma_1^{-1}\int_{\sigma_1}^{\frac{1}{2}} \frac{1}{\sin\left(\pi t\right)}\,dt \qquad (\sin(\pi t) \text{ is increasing on } [0,1/2]) \\
&\leq \frac{1}{\sin\left(\pi\sigma_1\right)} + \sigma_1^{-1}\sqrt{\frac{1}{2}-\sigma_1}\cdot\sqrt{\int_{\sigma_1}^{\frac{1}{2}} \frac{1}{\sin^2\left(\pi t\right)}\,dt} \qquad (\text{Cauchy-Shwarz inequality}) \\
&\leq \frac{1}{\sin\left(\pi\sigma_1\right)} + \frac{\sqrt{\cot(\pi\sigma_1)}}{\sigma_1\sqrt{2\pi}} \\
&= O\left(\left(\frac{N}{c_1}\right)^{\frac{3}{2}}\right). \qquad \left(\sigma_1 = \frac{c_1}{N},\ c_1 = o(N)\right)
\end{aligned} \tag{39}
$$

Substituting the estimate above back into (38), we find that

$$
\begin{aligned}
\left|z_{\sigma_1}^1(t)\right| &\leq \max_i |\alpha[i]|\cdot\frac{O(e^{-C_3c_1})}{\sqrt{N}} \sum_{0<i\sigma_1\leq\frac{1}{2}} \frac{1}{\sin\left(\pi i\sigma_1\right)} + \|n_{\sigma_1}(\cdot)\|_{L_\infty} \\
&= \max_i |\alpha[i]|\cdot O(Nc_1^{-\frac{3}{2}}e^{-C_3c_1}) + \|n_{\sigma_1}(\cdot)\|_{L_\infty} \\
&\leq \max_i |\alpha[i]|\cdot O(Ne^{-Cc_1}) + \|n_{\sigma_1}(\cdot)\|_{L_\infty} \qquad (c_1\to\infty) \\
&\leq 2\left\|n_{\sigma_1}(\cdot)\right\|_{L_\infty}, \qquad (c_1 = \Theta(\log N) \text{ with large enough lower bound})
\end{aligned} \tag{40}
$$

asymptotically. Let us simplify the noise term $\|n_{\sigma_1}(\cdot)\|_{L_\infty}$. Note that

$$
\begin{aligned}
&\|n_{\sigma_1}(\cdot)\|_{L_\infty} \\
&= \left\|\left(g_{\sigma_1,N}\circledast n\right)(\cdot)\right\|_{L_\infty} \qquad (\text{see } (36)) \\
&= \left\|\int_{\mathbb{I}} g_{\sigma_1,N}(t')\cdot n(t-t')\,dt'\right\|_{L_\infty} \\
&\leq \left\|g_{\sigma_1,N}(\cdot)\right\|_{L_1}\left\|n(\cdot)\right\|_{L_\infty} \qquad (\text{Holder inequality}) \\
&\leq \left\|g_{\sigma_1,N}(\cdot)\right\|_{L_2}\left\|n(\cdot)\right\|_{L_\infty} = \left\|n(\cdot)\right\|_{L_\infty}. \qquad (\text{Cauchy–Schwarz inequality, and Criterion 1})
\end{aligned} \tag{41}
$$

Overall, from (40), we conclude that

$$
\left|z_{\sigma_1}^1(t)\right| \leq 2\left\|n_{\sigma_1}(\cdot)\right\|_{L_\infty} \leq 2\left\|n(\cdot)\right\|_{L_\infty}, \qquad \text{if } \min_i d(t,\tau[i]) \geq \sigma_1, \tag{42}
$$

asymptotically. In words, $|z_{\sigma_1}^1(\cdot)|$ is small away from the impulses.

At impulses, on the contrary, $|z_{\sigma_1}^1(\cdot)|$ remains large as we argue next. Without loss of generality, consider the first impulse positioned at $\tau[1]$. We observe that

$$
\begin{aligned}
&\left|z^1_{\sigma_1}(\tau[1])\right| \\
&= \left|\sum_{i=1}^{K} \alpha[i]\cdot g_{\sigma_1,N}(\tau[1]\ominus\tau[i]) + n_{\sigma_1}(\tau[1])\right| \qquad \text{(see (38))} \\
&\geq |\alpha[1]|\cdot|g_{\sigma_1,N}(0)| - \max_i|\alpha[i]|\cdot\sum_{i=2}^{K}|g_{\sigma_1,N}(\tau[1]\ominus\tau[i])| - \|n_{\sigma_1}(\cdot)\|_{L_\infty} \\
&= |\alpha[1]|\cdot\Omega\left(\sqrt{\frac{N}{c}}\right) - 2\|n_{\sigma_1}(\cdot)\|_{L_\infty} \qquad \text{(Criterion 1, similar to (38))} \\
&\geq \min_j|\alpha[j]|\cdot\Omega\left(\sqrt{\frac{N}{c}}\right) - 2\|n(\cdot)\|_{L_\infty}. \qquad \text{(see (41))}
\end{aligned}
$$

Next, we introduce the dynamic range of the signal (namely, $\mathrm{dyn}(x_{\tau,\alpha})$) in order to simplify the expressions. More specifically, we continue by writing that

$$
\begin{aligned}
&\left|z^1_{\sigma_1}(\tau[1])\right| \\
&\geq \frac{\max_j|\alpha[j]|}{\mathrm{dyn}(x_{\tau,\alpha})}\cdot\Omega\left(\sqrt{\frac{N}{c}}\right) - 2\|n(\cdot)\|_{L_\infty} \qquad \text{(see (10))} \\
&\geq \frac{\max_j|\alpha[j]|}{\mathrm{dyn}(x_{\tau,\alpha})}\cdot\Omega\left(\sqrt{\frac{N}{c}}\right) - 2\|n(\cdot)\|_{L_\infty},
\end{aligned} \tag{43}
$$

asymptotically. In words, (43) states that $z^1_{\sigma_1}(\tau[i])$ is bounded away from zero (for every i). Put differently, for large enough N, there exists a constant $C_4 > 0$ such that

$$
\left|z^1_{\sigma_1}(\tau[i])\right| \geq C_4\cdot\frac{\|\alpha\|_\infty}{\mathrm{dyn}(x_{\tau,\alpha})}\sqrt{\frac{N}{c}} - 2\|n(\cdot)\|_{L_\infty}, \qquad i\in[1:K]. \tag{44}
$$

By comparing (42) and (44), we observe that if

$$
\frac{\|n(\cdot)\|_{L_\infty}}{\|\alpha\|_\infty} \leq \frac{C_4}{4}\cdot\frac{1}{\mathrm{dyn}(x_{\tau,\alpha})}\cdot\sqrt{\frac{N}{c}}, \tag{45}
$$

the lower bound is (44) does not exceed the upper bound in (42). All quantities $\|\alpha\|_\infty$, $\mathrm{dyn}(x_{\tau,\alpha})$, and $\|n(\cdot)\|_{L_\infty}$ are independent of c and N. Consequently, (45) is met asymptotically (i.e., for large enough N). As a result, $\tau^0[1]$ (where $|z^1_{\sigma_1}(\cdot)|$ achieves its maximum on $\mathbb{I}$) is within a radius σ_1 of the set τ, i.e.,

$$
\min_i d(\tau^0[1],\tau[i]) \leq \sigma_1 < \frac{1}{2}.
$$

Without loss of generality, suppose that $\tau[1]$ is the unique entry of τ that achieves the minimum above, i.e. $d(\tau^0[1],\tau[1]) = \min_i d(\tau^0[1],\tau[i])$. Indeed, the uniqueness is guaranteed because τ is asymptotically well-separated (see (35)). Then, according to (42), setting to zero a neighborhood of radius $2\sigma_1$ of $\tau^0[1]$ (to obtain $z^2_{\sigma_1}(\cdot)$) removes the *bump* located at $\tau[1]$. At the same time, since $\mathrm{sep}(\tau)\geq 4\sigma_1$ by (35), altering this neighborhood does not remove the bumps located at $\tau[i]$, $i>1$. Therefore, K repetitions of this process recovers every member of τ to a precision of σ_1. The algorithm terminates after K iterations (so that $\widetilde{K}=K$) because

$$
\|z^{K+1}_{\sigma_1}(\cdot)\|_{L_\infty} \leq 2\|n(\cdot)\|_{L_\infty} = \eta,
$$

asymptotically and according to (42). In other words, at this point, all the bumps have been removed and we have reached the noise/interference level. This completes the proof of Proposition 2.

6.2 Proof of Theorem 6 (Phase II)

At this point, we begin to study the performance of Algorithm II. *Stationarity* is a necessary (first-order) condition for a feasible point in $\mathbb{B}(\tau^0, \sigma_1)$ to be a local minimizer of Program (23). In a constrained program, a feasible point is stationary if the gradient of the objective function makes an acute angle with every feasible direction. To be concrete, we recall the definition of a stationary point [16] (slightly adjusted to match our settings).

Definition 10. [Stationary point] *In Program* (23), $\rho_s \in \mathbb{B}(\tau^0, \sigma_1)$ *is a stationary point if and only if*

$$\left\langle \frac{\partial F}{\partial \rho}(\rho_s), sign\left((\rho_s \ominus \rho) - \frac{1}{2}\right)\right\rangle \geq 0, \qquad \forall \rho \in \mathbb{B}(\tau^0, \sigma_1).$$

The entries of the sign vector above are $\{sign((\rho_s[i] \ominus \rho[i]) - \frac{1}{2})\}$, $i \in [1 : \widetilde{K}]$.

While not the focus of our analysis, one can establish that the gradient projection algorithm outlined in (27) (with appropriate step sizes $\{\delta^j\}$) always converges to a stationary point of Program (23). (Also confer [16, Theorem 5.4.6].)

Similarly, we prove next that the projected Newton's method in Algorithm II converges to a stationary point of Program (23). This claim depends on the following result adapted from [16, Theorem 5.5.2].

Proposition 11. [Convergence to a stationary point] *Any limit point of the sequence* $\{\tau^j\}_j$ *produced by Algorithm II is a stationary point of Program* (23) *if*

- *the gradient is Lipschitz continuous, i.e.,*

$$\left\| \frac{\partial F}{\partial \rho}(\rho_1) - \frac{\partial F}{\partial \rho}(\rho_2) \right\|_2 \leq L \cdot d\left(\rho_1, \rho_2\right), \qquad \forall \rho \in \mathbb{B}(\tau^0, \sigma_1),$$

  for some finite L*,*

- *the Hessian is positive definite on the feasible set, i.e.,*

$$\frac{\partial^2 F}{\partial \rho^2}(\rho) \succ 0, \qquad \forall \rho \in \mathbb{B}(\tau^0, \sigma_1),$$

- *both the spectral norm and the condition number of the Hessian are bounded on* $\mathbb{B}(\tau^0, \sigma_1)$*, and*

- *lastly,* $0 < \bar{\epsilon} \leq \epsilon^j < \sigma_1$ *for every* j *and for some* $\bar{\epsilon}$*.*

By (28), $\frac{\partial F}{\partial \rho}(\cdot)$ is continuous, and since $\mathbb{B}(\tau^0, \sigma_1)$ is compact, $\frac{\partial F}{\partial \rho}(\cdot)$ is Lipschitz contiuous too. In Appendix B, we establish that $\frac{\partial^2 F}{\partial \rho^2}(\cdot)$ is asymptotically positive definite on $\mathbb{B}(\tau^0, \sigma_1)$ (and moreover bounded from below by a positive factor of identity matrix) as long as

$$\frac{\|n(\cdot)\|_{L_2}}{\|\alpha\|_\infty} = \frac{O(1)}{\mathrm{dyn}\left(x_{\tau,\alpha}\right)^2},$$

with a small enough constant. Then, since the eigenvalues of a matrix are continuous functions of its entries, it follows that both spectral norm and condition number of the Hessian are bounded on $\mathbb{B}(\tau^0, \sigma_1)$. The last item in Proposition 11 holds by design (see Algorithm II). In summary, Proposition 11 is in force and any limit point of Algorithm II is a stationary point of Program (23).

Upon existence, let $\widetilde{\tau} \in \mathbb{I}^K$ denote one such limit point which, by Definition 10, satisfies

$$\left\langle \frac{\partial F}{\partial \rho}(\widetilde{\tau}), \operatorname{sign}\left((\widetilde{\tau} \ominus \rho) - \frac{1}{2}\right) \right\rangle \geq 0, \qquad \forall \rho \in \mathbb{B}(\tau^0, \sigma_1). \tag{46}$$

To control the distance of $\widetilde{\tau}$ from the true position vector τ, we upper-bound the above inner product as follows. See Appendix C for the proof.

Lemma 12. *For integer N and $0 < c_1 < c_2$ (both functions of N), let $\sigma_1 = \frac{c_1}{N}$ and $\sigma_2 = \frac{c_2}{N}$. Suppose that the kernel $g_{\sigma_2,N}(\cdot)$ satisfies Criteria 4 and 5 (with $\sigma = \sigma_2$). Suppose also that $2\sigma_1 \leq h(\sigma_2, N) \leq \sigma_2$ (see Criterion 5). Lastly, define $F(\cdot)$ as in (22), and recall the quantities involved there.*

Then, for every $\rho \in \mathbb{B}(\tau^0, \sigma_1)$, it holds asymptotically that

$$\begin{aligned}\left\langle \frac{\partial F}{\partial \rho}(\rho), sign\left((\rho \ominus \tau) - \frac{1}{2}\right) \right\rangle &= -\frac{\Omega(N)}{dyn(x_{\tau,\alpha})^2} \cdot \|\alpha\|_2^2 \cdot d(\rho, \tau) + O(e^{-Cc_2}) \cdot \|\alpha\|_2^2 \\ &\quad + O(N) \cdot \|n(\cdot)\|_{L_2} \|\alpha\|_2 + O(N) \cdot \|n(\cdot)\|_{L_2}^2,\end{aligned}$$

when $c_1, c_2, N \to \infty$ and $c_1, c_2 = \Theta(\log N)$ (with a large enough lower bound).

We are now ready to complete the proof of Theorem 6. Since $\tau \in \mathbb{B}(\tau^0, \sigma_1)$ too, in light of (46) and Lemma 12, we can write that

$$\begin{aligned}0 \leq \left\langle \frac{\partial F}{\partial \rho}(\widetilde{\tau}), \operatorname{sign}\left((\widetilde{\tau} \ominus \tau) - \frac{1}{2}\right) \right\rangle &= -\frac{\Omega(N)}{\operatorname{dyn}(x_{\tau,\alpha})^2} \cdot \|\alpha\|_2^2 \cdot d(\widetilde{\tau}, \tau) + O(e^{-Cc_2}) \cdot \|\alpha\|_2^2 \\ &\quad + O(N) \cdot \|n(\cdot)\|_{L_2} \|\alpha\|_2 + O(N) \cdot \|n(\cdot)\|_{L_2}^2,\end{aligned}$$

which simplifies to

$$\begin{aligned}\frac{d(\widetilde{\tau}, \tau)}{\operatorname{dyn}(x_{\tau,\alpha})^2} &= O\left(e^{-Cc_2} + \frac{\|n(\cdot)\|_{L_2}}{\|\alpha\|_2} + \frac{\|n(\cdot)\|_{L_2}^2}{\|\alpha\|_2^2}\right) \\ &= O\left(e^{-Cc_2} + \frac{\|n(\cdot)\|_{L_2}}{\|\alpha\|_2}\right), \qquad \text{if } \|n(\cdot)\|_{L_2} \leq \|\alpha\|_2,\end{aligned}$$

asymptotically. This completes the proof of Theorem 6 since we already know that

$$\widetilde{\tau}, \tau \in \mathbb{B}(\tau^0, \sigma_1) \Longrightarrow d(\widetilde{\tau}, \tau) \leq d(\widetilde{\tau}, \tau^0) + d(\tau^0, \tau) \leq 2\sigma_1,$$

under Proposition 2.

Acknowledgments

AE acknowledges Ben Adcock, Aditya Viswanathan, and Anne Gelb for pointing out the possible connection between our work and [5]. Part of this research was conducted when AE was a graduate fellow at the Statistical and Applied Mathematical Sciences Institute (SAMSI) and later a visitor at the Institute for Computational and Experimental Research in Mathematics (ICERM). AE is grateful for their hospitality and kindness.

References

[1] J.M. Azais, Y. De Castro, and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

[2] E.J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[3] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[4] E.J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[5] D. Cochran, A. Gelb, and Y. Wang. Edge detection from truncated Fourier data using spectral mollifiers. *Advances in computational mathematics*, pages 1–26, 2013.

[6] M.A. Davenport and M.B. Wakin. Compressive sensing of analog signals using discrete prolate spheroidal sequences. *Applied and Computational Harmonic Analysis*, 33(3):438–472, 2012.

[7] L. Demanet, D. Needell, and N. Nguyen. Super-resolution via superset selection and pruning. *arXiv preprint arXiv:1302.6288*, 2013.

[8] V. Duval and G. Peyre. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, pages 1–41, 2015.

[9] A. Eftekhari, J. Romberg, and M.B. Wakin. Matched filtering from limited frequency samples. *IEEE Transactions on Information Theory*, 59(6):3475–3496, 2013.

[10] A. Eftekhari and M.B. Wakin. Greed is super: A new iterative method for super-resolution. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2013.

[11] A. Eftekhari and M.B. Wakin. Supplementary material for "Greed is super: A fast algorithm for super-resolution". Technical report, Colorado School of Mines, 2015.

[12] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM Journal on Imaging Sciences*, 5(1):179–202, 2012.

[13] C. Fernandez-Granda. Support detection in super-resolution. *arXiv preprint arXiv:1302.3921*, 2013.

[14] K. Fyhn, H. Dadkhahi, and M.F. Duarte. Spectral compressive sensing with polar interpolation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[15] J.A. Hogan and J.D. Lakey. *Duration and bandwidth limiting: Prolate functions, sampling, and applications.* Applied and Numerical Harmonic Analysis. Birkhauser Boston, 2012.

[16] C.T. Kelley. *Iterative methods for optimization.* Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics, 1999.

[17] R.H. Keshavan and S. Oh. A gradient descent algorithm on the Grassman manifold for matrix completion. *arXiv preprint arXiv:0910.5260*, 2009.

[18] D.E. Knuth. Big omicron and big omega and big theta. *ACM Sigact News*, 8(2):18–24, 1976.

[19] W. Liao and A. Fannjiang. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 2014.

[20] S. G Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[21] A. Meister. *Deconvolution problems in nonparametric statistics.* Lecture Notes in Statistics. Springer Berlin Heidelberg, 2009.

[22] A. Osipov, V. Rokhlin, and H. Xiao. *Prolate spheroidal wave functions of order zero: Mathematical tools for bandlimited approximation.* Applied Mathematical Sciences. Springer, 2013.

[23] D. Slepian. Some asymptotic expansions for prolate spheroidal wave functions. *Journal of Mathematical Physics*, 44(2):99–140, 1965.

[24] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty V. *Bell Systems Technical Journal*, 57(5):1371–1429, 1978.

[25] G. Tang, B.N. Bhaskar, and B. Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2015.

[26] G. Tang, B.N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.

[27] D.J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

A Toolbox

This section collects a number of results which are frequently invoked in the rest of the appendix.

In what follows, with integer N and $c = c(N) > 0$, we assume that $\sigma = \frac{c}{N}$, and consider a kernel $g_{\sigma,N}(\cdot) = g(\cdot;\sigma,N)$ that satisfies Criteria 4 and 5.

Lemma 13. *For integer N and $c = c(N) > 0$, let $\sigma = \frac{c}{N}$. Consider a kernel $g_{\sigma,N}(\cdot)$ that satisfies Criterion 4. Fix $i \in [1:K]$, and $\rho \in \mathbb{I}^K$ with distinct entries. Then, it holds asymptotically that*

$$\sum_{j\in[1:K]\backslash\{i\}} |\langle g_{\sigma,N}(t\ominus\rho[i]), g_{\sigma,N}(t\ominus\rho[j])\rangle| = O(e^{-Cc}), \tag{47}$$

$$\sum_{j\in[1:K]\backslash\{i\}} \left|\left\langle g_{\sigma,N}(t\ominus\rho[i]), g'_{\sigma,N}(t\ominus\rho[j])\right\rangle\right| = O(e^{-Cc}), \tag{48}$$

$$\sum_{j\in[1:K]\backslash\{i\}} \left|\left\langle g'_{\sigma,N}(t\ominus\rho[i]), g'_{\sigma,N}(t\ominus\rho[j])\right\rangle\right| = O(e^{-Cc}), \tag{49}$$

when $c, N \to \infty$ and $c = \Theta(\log N)$ (with a large enough lower bound).

Proof. These inequalities are direct consequences of Criterion 4. Indeed, since the entries of ρ are distinct, $\text{sep}(\rho) \geq 2\sigma$ asymptotically (i.e., for large enough N). Then, to prove (47), we write that

$$\begin{aligned}
&\sum_{j\in[1:K]\backslash\{i\}} |\langle g_{\sigma,N}(t\ominus\rho[i]), g_{\sigma,N}(t\ominus\rho[j])\rangle| \\
&= O\left(\frac{e^{-C_3 c}}{N}\right) \sum_{j\in[1:K]\backslash\{i\}} \frac{1}{|\sin\left(\pi\cdot d(\rho[i],\rho[j])\right)|} \qquad \text{(see Criterion 4)} \\
&= O\left(\frac{e^{-C_3 c}}{N}\right) \sum_{0<l\cdot 2\sigma\leq\frac{1}{2}} \frac{1}{\sin\left(\pi\cdot l\cdot 2\sigma\right)} \qquad (\text{sep}(\tau)\geq 2\sigma,\ \text{asymptotically}) \\
&= O\left(\frac{e^{-C_3 c}}{N}\right)\left(\frac{N}{c}\right)^{\frac{3}{2}} \qquad \text{(similar to (39))} \\
&= O\left(N^{\frac{1}{2}} e^{-Cc}\right) \qquad (c\to\infty) \\
&= O\left(e^{-Cc}\right). \qquad (c=\Theta(\log N))
\end{aligned}$$

The next inequalities in Lemma 13 are proved similarly and we omit the details here. $\square$

Lemma 14. *For integer N and $c = c(N) > 0$, let $\sigma = \frac{c}{N}$. Consider a kernel $g_{\sigma,N}(\cdot)$ that satisfies Criterion 4. Fix $i \in [1:K]$, and $\rho_1, \rho_2 \in \mathbb{I}^K$. Suppose that $\rho_2[j] \neq \rho_1[i]$, for every $j \neq i$. Then, it holds asymptotically that*

$$\sum_{j\in[1:K]\backslash\{i\}} |\langle g_{\sigma,N}(t\ominus\rho_1[i]), g_{\sigma,N}(t\ominus\rho_2[j])\rangle| = O(e^{-Cc}), \tag{50}$$

$$\sum_{j\in[1:K]\backslash\{i\}} \left|\left\langle g_{\sigma,N}(t\ominus\rho_1[i]), g'_{\sigma,N}(t\ominus\rho_2[j])\right\rangle\right| = O(e^{-Cc}), \tag{51}$$

when $c, N \to \infty$ and $c = \Theta(\log N)$ (with a large enough lower bound).

Proof. Note that, by hypothesis, the vector formed from $\{\rho_1[i]\} \cup \{\rho_2[j]\}_{j\neq i}$ has distinct entries to which we can apply Lemma 13. This completes the proof of Lemma 14. $\square$

A few more technical lemmas are in order. In what follows, $G_\rho = G_{\rho,\sigma,N} \in \mathbb{C}^{N\times K}$ is defined similar to (14) for $\rho \in \mathbb{I}^K$, $\sigma < \frac{1}{2}$, and integer N. Among other things, the next result states that the columns of G_ρ are nearly orthonormal as long as ρ is well-separated.

Lemma 15. *For integer N and $c = c(N) > 0$, let $\sigma = \frac{c}{N}$. Consider a kernel $g_{\sigma,N}(\cdot)$ that satisfies Criterion 4. Fix $\rho \in \mathbb{I}^K$ with distinct entries and recall (14). Then, it holds asymptotically that*

$$\left\|I_K - G_\rho^* G_\rho\right\| = O(e^{-Cc}), \tag{52}$$

$$\|G_\rho\| \leq 1 + O(e^{-Cc}), \tag{53}$$

$$\left\|\left(G_\rho^* G_\rho\right)^{-1}\right\| \leq 1 + O(e^{-Cc}), \tag{54}$$

$$\left\|G_\rho^\dagger\right\| \leq 1 + O(e^{-Cc}), \tag{55}$$

$$\left\|G_\rho^* L G_\rho\right\| = O(e^{-Cc}), \tag{56}$$

$$\left\|\left\|g'_{\sigma,N}(\cdot)\right\|_{L_2}^2 \cdot I_K - G_\rho^* L^* L G_\rho\right\| = O(e^{-Cc}), \tag{57}$$

as $c, N \to \infty$ and $c = \Theta(\log N)$ (with a sufficiently large lower bound). Above, $A^\dagger$ is the pseudo-inverse of A, and $\|A\|$ returns its spectral norm of A. Entries of the diagonal matrix $L \in \mathbb{C}^{N\times N}$ are specified as $L[l,l] = i2\pi l$, $l \in \mathbb{F}$. The inverse of $G_\rho^* G_\rho$ exists, so that (54) and (55) are well-defined.

Moreover, suppose that $\rho_1, \rho_2 \in \mathbb{I}^K$ satisfy $\rho_1[i] \neq \rho_2[j]$ for all $i \neq j$. Then, it holds asymptotically that

$$\left\|G_{\rho_1}^* G_{\rho_2} - M_{\rho_1,\rho_2}\right\| = O(e^{-Cc}), \tag{58}$$

where the entries of diagonal matrix $M_{\rho_1,\rho_2} \in \mathbb{R}^{K\times K}$ are specified as

$$M_{\rho_1,\rho_2}[i,i] = \langle g_{\sigma,N}(t \ominus \rho_1[i]), g_{\sigma,N}(t \ominus \rho_2[i])\rangle, \qquad i \in [1:K]. \tag{59}$$

It also holds asymptotically that

$$\left\|G_{\rho_1}^* L G_{\rho_2} - M_{\rho_1,\rho_2}^d\right\| = O(e^{-Cc}), \tag{60}$$

where the diagonal matrix $M_{\rho_1,\rho_2}^d \in \mathbb{R}^{K\times K}$ is defined with

$$M_{\rho_1,\rho_2}^d[i,i] = \left\langle g_{\sigma,N}(t \ominus \rho_1[i]), g'_{\sigma,N}(t \ominus \rho_2[i])\right\rangle, \qquad i \in [1:K]. \tag{61}$$

In addition,

$$\|G_{\rho_1} - G_{\rho_2}\| \leq \sqrt{2}\,\|I_K - M_{\rho_1,\rho_2}\|^{\frac{1}{2}} + O(e^{-Cc}), \tag{62}$$

$$\left\|G_{\rho_1}^\dagger - G_{\rho_2}^\dagger\right\| = O(1)\,\|I_K - M_{\rho_1,\rho_2}\|^{\frac{1}{2}} + O(e^{-Cc}). \tag{63}$$

Proof. Because $\|g_{\sigma,N}(\cdot)\|_{L_2} = 1$ by Criterion 4, the diagonal entries of $G_\rho^* G_\rho$ equal to one, and in fact

$$\left(I_K - G_\rho^* G_\rho\right)[i,j] = \begin{cases} 0 & i = j, \\ -\langle g_{\sigma,N}(t \ominus \rho[i]), g_{\sigma,N}(t \ominus \rho[j])\rangle & i \neq j, \end{cases} \tag{64}$$

where I_K is the $K \times K$ identity matrix. Let $\lambda_l(A)$ return the lth eigenvalue of a square matrix A. Then, using the Gershgorin disc theorem, we can write that

$$\begin{aligned} \left\|I_K - G_\rho^* G_\rho\right\| &= \max_{l\in[1:K]} \left|\lambda_l\left(I_K - G_\rho^* G_\rho\right)\right| \\ &\leq \max_{i\in[1:K]} \sum_{j\neq i} |\langle g_{\sigma,N}(t \ominus \rho[i]), g_{\sigma,N}(t \ominus \rho[j])\rangle| \qquad \text{(see (64))} \\ &= O(e^{-Cc}). \qquad \text{(see (47))} \end{aligned} \tag{65}$$

This establishes (52). It also follows that

$$\begin{aligned} \|G_\rho\|^2 - 1 &= \left\|G_\rho^* G_\rho\right\| - 1 \\ &\leq \left\|I_K - G_\rho^* G_\rho\right\| \\ &= O(e^{-Cc}), \end{aligned}$$

which implies (53). Similarly, letting $\sigma_i(A)$ be the lth singular value of a matrix A, we can write that

$$\begin{aligned} \min_{i\in[1:K]} \sigma_i\left(G_\rho^* G_\rho\right) - 1 &= \min_{i\in[1:K]} \lambda_i\left(G_\rho^* G_\rho\right) - 1 \\ &= -\max_{i\in[1:K]} \lambda_i\left(I_K - G_\rho^* G_\rho\right) \\ &\geq -\max_{i\in[1:K]} \left|\lambda_i\left(I_K - G_\rho^* G_\rho\right)\right| \\ &= -\left\|I_K - G_\rho^* G_\rho\right\| \\ &= -O(e^{-Cc}). \qquad \text{(see (52))} \end{aligned}$$

It immediately follows that

$$\left\| \left(G_\rho^* G_\rho\right)^{-1} \right\| = \frac{1}{\min_i \sigma_i \left(G_\rho^* G_\rho\right)} \le \frac{1}{1 - O(e^{-Cc})} = 1 + O(e^{-Cc}),$$

as claimed in (54). Additionally, (55) follows directly from (53) and (54). We next observe that

$$(G_\rho^* L G_\rho)[i,j] = \begin{cases} 0 & i = j, \\ \left\langle g_{\sigma,N}(t \ominus \rho[i]), g'_{\sigma,N}(t \ominus \rho[j]) \right\rangle & i \ne j, \end{cases}$$

where we used the fact that $\left\langle g_{\sigma,N}(t \ominus \rho[i]), g'_{\sigma,N}(t \ominus \rho[i]) \right\rangle = 0$ because $g_{\sigma,N}(t)$ is symmetric about $t = \frac{1}{2}$ (and hence $g'_{\sigma,N}(\cdot)$ is anti-symmetric about $t = \frac{1}{2}$). Using the Gershgorin disc theorem once more, it follows that

$$\begin{aligned} \left\| G_\rho^* L G_\rho \right\| &= \max_{l \in [1:K]} \left| \lambda_l \left(G_\rho^* L G_\rho \right) \right| \\ &\le \max_{i \in [1:K]} \sum_{j \ne i} \left| (G_\rho^* L G_\rho)[i,j] \right| \\ &= \max_{i \in [1:K]} \sum_{j \ne i} \left| \left\langle g_{\sigma,N}(t \ominus \rho[i]), g'_{\sigma,N}(t \ominus \rho[j]) \right\rangle \right| \qquad \text{(see (14))} \\ &= O(e^{-Cc}), \qquad \text{(see (48))} \end{aligned}$$

where the last line uses (48). This establishes (56). The proof of (57) is similar to that of (52) and is omitted here.

Next, by the definition of M_{ρ_1,ρ_2} in (59), it holds that

$$\left(G_{\rho_1}^* G_{\rho_2} - M_{\rho_1,\rho_2} \right)[i,j] = \begin{cases} 0 & i = j, \\ \langle g_{\sigma,N}(t \ominus \rho_1[i]), g_{\sigma,N}(t \ominus \rho_2[j]) \rangle & i \ne j. \end{cases}$$

We can therefore write that

$$\begin{aligned} & \left\| G_{\rho_1}^* G_{\rho_2} - M_{\rho_1,\rho_2} \right\| \\ & \le \max \left[\left\| G_{\rho_1}^* G_{\rho_2} - M_{\rho_1,\rho_2} \right\|_{1,1}, \left\| G_{\rho_1}^* G_{\rho_2} - M_{\rho_1,\rho_2} \right\|_{\infty,\infty} \right] \quad (\|A\| \le \max\left[\|A\|_{1,1}, \|A\|_{\infty,\infty}\right]) \\ & = \max \left[\max_{i \in [1:K]} \sum_{j \ne i} |\langle g_{\sigma,N}(t \ominus \rho_1[i]), g_{\sigma,N}(t \ominus \rho_2[j]) \rangle| \right. \\ & \qquad \left. , \max_{j \in [1:K]} \sum_{i \ne j} |\langle g_{\sigma,N}(t \ominus \rho_1[i]), g_{\sigma,N}(t \ominus \rho_2[j]) \rangle| \right] \\ & = O(e^{-Cc}), \qquad \text{(see (50))} \end{aligned}$$

where $\|A\|_{1,1}$ and $\|A\|_{\infty,\infty}$ are $\ell_1 \to \ell_1$ and $\ell_\infty \to \ell_\infty$ operator norms of matrix A. This proves (58). Similarly, recalling (61), we note that

$$\left(G_{\rho_1}^* L G_{\rho_2} - M_{\rho_1,\rho_2}^d \right)[i,j] = \begin{cases} 0 & i = j, \\ \left\langle g_{\sigma,N}(t \ominus \rho_1[i]), g'_{\sigma,N}(t \ominus \rho_2[j]) \right\rangle & i \ne j, \end{cases}$$

from which it follows that

$$
\begin{aligned}
&\left\|G^*_{\rho_1} L G_{\rho_2} - M^d_{\rho_1,\rho_2}\right\| \\
&\le \max\left[\left\|G^*_{\rho_1} L G_{\rho_2} - M^d_{\rho_1,\rho_2}\right\|_{1,1}, \left\|G^*_{\rho_1} L G_{\rho_2} - M^d_{\rho_1,\rho_2}\right\|_{\infty,\infty}\right] \ \left(\|A\| \le \max\left[\|A\|_{1,1}, \|A\|_{\infty,\infty}\right]\right) \\
&= \max\Big[\max_{i\in[1:K]} \sum_{j\ne i} \left|\left\langle g_{\sigma,N}(t\ominus\rho_1[i]), g'_{\sigma,N}(t\ominus\rho_2[j])\right\rangle\right| \\
&\qquad , \max_{j\in[1:K]} \sum_{i\ne j} \left|\left\langle g_{\sigma,N}(t\ominus\rho_1[i]), g'_{\sigma,N}(t\ominus\rho_2[j])\right\rangle\right|\Big] \\
&= O(e^{-Cc}). \qquad \text{(see (51))}
\end{aligned}
$$

This establishes (60). To prove (62), we note that

$$
\begin{aligned}
&\|G_{\rho_1} - G_{\rho_2}\|^2 \\
&= \left\|\left(G_{\rho_1} - G_{\rho_2}\right)^* \left(G_{\rho_1} - G_{\rho_2}\right)\right\| \\
&= \left\|G^*_{\rho_1} G_{\rho_1} + G^*_{\rho_2} G_{\rho_2} - G^*_{\rho_1} G_{\rho_2} - G^*_{\rho_2} G_{\rho_1}\right\| \\
&= \Big\|2I_K - 2M_{\rho_1,\rho_2} - \left(I_K - G^*_{\rho_1} G_{\rho_1}\right) - \left(I_K - G^*_{\rho_2} G_{\rho_2}\right) \\
&\qquad - \left(G^*_{\rho_1} G_{\rho_2} - M_{\rho_1,\rho_2}\right) - \left(G^*_{\rho_2} G_{\rho_1} - M_{\rho_1,\rho_2}\right)\Big\| \\
&\le 2\left\|I_K - M_{\rho_1,\rho_2}\right\| + \left\|I_K - G^*_{\rho_1} G_{\rho_1}\right\| + \left\|I_K - G^*_{\rho_2} G_{\rho_2}\right\| \\
&\qquad + 2\left\|G^*_{\rho_1} G_{\rho_2} - M_{\rho_1,\rho_2}\right\| \quad \text{(see (59))} \\
&= 2\left\|I_K - M_{\rho_1,\rho_2}\right\| + O(e^{-Cc}), \qquad \text{(see (52) and (58))}.
\end{aligned}
$$

Lastly, to prove (63), we write that

$$
\begin{aligned}
&\left\|G^\dagger_{\rho_1} - G^\dagger_{\rho_2}\right\| \\
&= \left\|\overbrace{\left(G^*_{\rho_1} G_{\rho_1}\right)^{-1}}^{A} \cdot \overbrace{G_{\rho_1}}^{B} - \overbrace{\left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}}^{C} \cdot \overbrace{G_{\rho_2}}^{D}\right\| \\
&\le \left\|\left[\left(G^*_{\rho_1} G_{\rho_1}\right)^{-1} - \left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}\right] G_{\rho_1}\right\| + \left\|\left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}\left[G_{\rho_1} - G_{\rho_2}\right]\right\| \\
&\le \left\|\left(G^*_{\rho_1} G_{\rho_1}\right)^{-1} - \left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}\right\| \cdot \|G_{\rho_1}\| + \left\|\left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}\right\| \cdot \|G_{\rho_1} - G_{\rho_2}\| \\
&\le \left\|\left(G^*_{\rho_1} G_{\rho_1}\right)^{-1}\right\| \cdot \left\|G^*_{\rho_1} G_{\rho_1} - G^*_{\rho_2} G_{\rho_2}\right\| \cdot \left\|\left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}\right\| \cdot \|G_{\rho_1}\| + \left\|\left(G^*_{\rho_2} G_{\rho_2}\right)^{-1}\right\| \\
&\qquad \cdot \|G_{\rho_1} - G_{\rho_2}\| \\
&= O(1)\left\|\overbrace{G^*_{\rho_1}}^{A} \cdot \overbrace{G_{\rho_1}}^{B} - \overbrace{G^*_{\rho_2}}^{C} \cdot \overbrace{G_{\rho_2}}^{D}\right\| + O(1)\|G_{\rho_1} - G_{\rho_2}\| \qquad \text{(see (53) and (54))} \\
&\le O(1)\|G_{\rho_1} - G_{\rho_2}\| \cdot \max\left[\|G_{\rho_1}\|, \|G_{\rho_2}\|\right] + O(1)\|G_{\rho_1} - G_{\rho_2}\| \\
&= O(1)\|G_{\rho_1} - G_{\rho_2}\| + O(1)\|G_{\rho_1} - G_{\rho_2}\| \qquad \text{(see (53))} \\
&= O(1)\left(\sqrt{2\left\|I_K - M_{\rho_1,\rho_2}\right\|} + O(e^{-Cc})\right) \qquad \text{(see (62))} \\
&= O(1)\left\|I_K - M_{\rho_1,\rho_2}\right\|^{\frac{1}{2}} + O(e^{-Cc}).
\end{aligned}
$$

Above, we twice used the identity $AB - CD = (A - C)B + C(B - D)$ for conformal matrices A, B, C, D. The fifth line owes itself to the identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ for (conformal and invertible) matrices A, B.

This concludes the proof of Lemma 15. $\square$

If the entries of $\rho \in \mathbb{I}^K$ are distinct, $G_\rho \in \mathbb{C}^{N\times K}$ has nearly orthonormal columns asymptotically (by Lemma 15), and it holds that $\|G_\rho\beta\|_2 \approx \|\beta\|_2$ for any $\beta \in \mathbb{C}^K$. This is recorded next.

Lemma 16. *For integer N and $c = c(N) > 0$, let $\sigma = \frac{c}{N}$. Consider a kernel $g_{\sigma,N}(\cdot)$ that satisfies Criterion 4. Fix $\rho \in \mathbb{I}^K$ with distinct entries, and $\beta \in \mathbb{R}^K$, and recall (14). It then holds asymptotically that*

$$\left(1 - O(e^{-Cc})\right)\|\beta\|_2^2 \le \|G_\rho\beta\|_2^2 \le \left(1 + O(e^{-Cc})\right)\|\beta\|_2^2,$$

$$\left(1 - O(e^{-Cc})\right)\|\beta\|_2^2 \le \frac{\|LG_\rho\beta\|_2^2}{\left\|g'_{\sigma,N}(\cdot)\right\|_{L_2}^2} \le \left(1 + O(e^{-Cc})\right)\|\beta\|_2^2,$$

when $c, N \to \infty$ and $c = \Theta(\log N)$ (with a large enough lower bound).

Proof. This is a direct consequence of Lemma 15. Indeed, it holds asymptotically that

$$\begin{aligned}
\left|\|G_\rho\beta\|_2^2 - \|\beta\|_2^2\right| &= \left|\beta^*\left(G_\rho^*G_\rho - I_K\right)\beta\right| \\
&\le \left\|G_\rho^*G_\rho - I_K\right\|\|\beta\|_2^2 \\
&= O(e^{-Cc})\|\beta\|_2^2. \qquad \text{(see (52))}
\end{aligned}$$

The second claim is proved similarly using (57). This completes the proof of Lemma 16. $\square$

We close this section with the following auxiliary result that approximates certain projection matrices with simpler quantities.

Lemma 17. *For integer N and $c = c(N) > 0$, let $\sigma = \frac{c}{N}$. Consider a kernel $g_{\sigma,N}(\cdot)$ that satisfies Criterion 4. Consider a vector $u \in \mathbb{R}^K$ and let $U = diag(u) \in \mathbb{R}^{K\times K}$ be the diagonal matrix formed from u. Suppose that $\rho \in \mathbb{I}^K$ has distinct entries and set $\mathcal{P}_{\rho,U} := G_\rho U G_\rho^\dagger \in \mathbb{C}^{N\times N}$ (after recalling (14)).*[15] *Then, it holds asymptotically that*

$$\left\|\mathcal{P}_{\rho,U} - G_\rho U G_\rho^*\right\| = O(e^{-Cc})\|u\|_\infty, \tag{66}$$

$$\|\mathcal{P}_{\rho,U}\| \le 2\|u\|_\infty, \tag{67}$$

when $c, N \to \infty$ and $c = \Theta(\log N)$ (with a large enough lower bound).

Furthermore, suppose that $\rho_1, \rho_2 \in \mathbb{I}^K$ both have distinct entries and $\rho_1[i] \ne \rho_2[j]$ when $i \ne j$. Then, for any $\beta \in \mathbb{R}^K$, we asymptotically have that

$$\|\mathcal{P}_{\rho_2,U}G_{\rho_1}\beta - G_{\rho_2}UM_{\rho_1,\rho_2}\beta\|_2 = O(e^{-Cc})\|u\|_\infty\|G_{\rho_1}\beta\|_2, \tag{68}$$

with $M_{\rho_1,\rho_2} \in \mathbb{R}^{K\times K}$ defined as in (59).

[15] In particular, when $U = I_K$, $\mathcal{P}_{\rho,I_K}$ is the orthogonal projection onto $\text{span}(G_\rho)$.

Proof. We show that $\mathcal{P}_{\rho,U}$ can be well approximated with $G_\rho U G_\rho^*$, and do so by bounding $\|\mathcal{P}_{\rho,U} - G_\rho U G_\rho^*\|$ next. We use the fact that $G_\rho \in \mathbb{C}^{N\times K}$ has nearly orthonormal columns (thanks to the distinct entries of ρ). Asymptotically, it holds that

$$\begin{aligned}
\left\|\mathcal{P}_{\rho,U} - G_\rho U G_\rho^*\right\| &= \left\|G_\rho U \left(G_\rho^* G_\rho\right)^{-1} G_\rho^* - G_\rho U G_\rho^*\right\| \\
&\le \|G_\rho\| \, \|U\| \left\|\left(G_\rho^* G_\rho\right)^{-1} - I_K\right\| \|G_\rho\| \\
&= \|G_\rho\|^2 \, \|u\|_\infty \left\|\left(G_\rho^* G_\rho\right)^{-1} - I_K\right\| \qquad (U = \operatorname{diag}(u)) \\
&\le \|G_\rho\|^2 \, \|u\|_\infty \left\|\left(G_\rho^* G_\rho\right)^{-1}\right\| \left\|I_K - G_\rho^* G_\rho\right\| \\
&= O(e^{-Cc})\|u\|_\infty. \qquad \text{(see Lemma 15)}
\end{aligned} \tag{69}$$

This proves (66). Also, (67) is proved by noting that

$$\begin{aligned}
\|\mathcal{P}_{\rho,U}\| &\le \left\|G_\rho U G_\rho^*\right\| + \left\|\mathcal{P}_{\rho,U} - G_\rho U G_\rho^*\right\| \\
&\le \|G_\rho\|^2 \, \|u\|_\infty + O(e^{-Cc})\|u\|_\infty \qquad \text{(see (69))} \\
&\le 2\|u\|_\infty, \qquad \text{(see (53))}
\end{aligned}$$

asymptotically. Lastly, using the just-established (66) and (67), we prove (68) as follows. (We will use the triangle inequality and basic manipulations, and also the fact that the spectral norm of a diagonal matrix equals its maximum entry.) Asymptotically, it holds that

$$\begin{aligned}
&\|\mathcal{P}_{\rho_2,U} G_{\rho_1}\beta - G_{\rho_2} U M_{\rho_1,\rho_2}\beta\|_2 \\
&\le \left\|\mathcal{P}_{\rho_2,U} G_{\rho_1}\beta - G_{\rho_2} U G_{\rho_2}^* G_{\rho_1}\beta\right\|_2 + \left\|G_{\rho_2} U G_{\rho_2}^* G_{\rho_1}\beta - G_{\rho_2} U M_{\rho_1,\rho_2}\beta\right\|_2 \\
&\le \left\|\mathcal{P}_{\rho_2,U} - G_{\rho_2} U G_{\rho_2}^*\right\| \|G_{\rho_1}\beta\|_2 + \|G_{\rho_2}\| \|U\| \left\|G_{\rho_1}^* G_{\rho_2} - M_{\rho_1,\rho_2}\right\| \|\beta\|_2 \\
&= O(e^{-Cc})\|u\|_\infty \|G_{\rho_1}\beta\|_2,
\end{aligned}$$

where we also used Lemmas 15 and 16. This proves (68) and completes the proof of Lemma 17. $\square$

Lemma 18. *For integer N and $c = c(N) > 0$, let $\sigma = \frac{c}{N}$. Consider a kernel $g_{\sigma,N}(\cdot)$ that satisfies Criterion 4. Suppose that $\rho_1, \rho_2 \in \mathbb{I}^K$ satisfy $\rho_1[i] \neq \rho_2[j]$ for all $i \neq j$. Recall (14), and for vectors $\alpha \in \mathbb{R}^K$ and $\widehat{n} \in \mathbb{C}^N$, set*[16]

$$\beta_{\rho_2} = G_{\rho_2}^\dagger (G_{\rho_1}\alpha + \widehat{n}).$$

Then, it holds asymptotically that

$$\|\beta_{\rho_2} - \alpha\|_\infty = O(1) \cdot \left(\sqrt{K}\, \|I_K - M_{\rho_1,\rho_2}\|^{\frac{1}{2}} + e^{-Cc}\right) \|\alpha\|_\infty + O(1) \cdot \|\widehat{n}\|_2 \,,$$

$$\|\beta_{\rho_2} - \beta_{\rho_1}\|_\infty = O(1) \cdot \left(\|I_K - M_{\rho_1,\rho_2}\|^{\frac{1}{2}} + e^{-Cc}\right) \cdot \left(\sqrt{K}\|\alpha\|_\infty + \|\widehat{n}\|_2\right),$$

when $c, N \to \infty$ and $c = \Theta(\log N)$ (with a large enough lower bound).

[16] Dependence of β_{ρ_2} on other parameters (particularly, ρ_1) is suppressed for convenience.

Proof. Note that

$$
\begin{aligned}
&\|\beta_{\rho_2}-\alpha\|_\infty \\
&\le \left\|G_{\rho_2}^\dagger G_{\rho_1}\alpha-\alpha\right\|_\infty + \left\|G_{\rho_2}^\dagger \widehat{n}\right\|_\infty \qquad \left(\beta_{\rho_2}=G_{\rho_2}^\dagger\left(G_{\rho_1}\alpha+\widehat{n}\right)\right) \\
&= \left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)G_{\rho_1}\alpha\right\|_\infty + \left\|G_{\rho_2}^\dagger \widehat{n}\right\|_\infty \\
&\le \left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)G_{\rho_1}\right\|_{\infty\to\infty}\|\alpha\|_\infty + \left\|G_{\rho_2}^\dagger \widehat{n}\right\|_2 \\
&\le \sqrt{K}\left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)G_{\rho_1}\right\|\cdot\|\alpha\|_\infty + \left\|G_{\rho_2}^\dagger \widehat{n}\right\|_2 \qquad \left(\|A\|_{\infty\to\infty}\le\sqrt{K}\cdot\|A\|, \quad A\in\mathbb{C}^{N\times K}\right) \\
&\le \sqrt{K}\left\|G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right\|\|G_{\rho_1}\|\cdot\|\alpha\|_\infty + \left\|G_{\rho_1}^\dagger\right\|\|\widehat{n}\|_2 \\
&= O(1)\left(\sqrt{K}\,\|I_K-M_{\rho_1,\rho_2}\|^{\frac{1}{2}}+e^{-Cc}\right)\cdot\|\alpha\|_\infty \\
&\qquad + O(1)\,\|\widehat{n}\|_2\,. \qquad (\text{Lemma 15 and } c=\Theta(\log N)).
\end{aligned}
$$

The last line above requires the lower bound in $c=\Theta(\log N)$ to be sufficiently large. Similarly,

$$
\begin{aligned}
&\|\beta_{\rho_2}-\beta_{\rho_1}\|_\infty \\
&\le \left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)G_{\rho_1}\alpha\right\|_\infty + \left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)\widehat{n}\right\|_\infty \qquad \left(\beta_{\rho_1}=G_{\rho_1}^\dagger\left(G_{\rho_1}\alpha+\widehat{n}\right)\right) \\
&\le \left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)G_{\rho_1}\right\|_{\infty\to\infty}\|\alpha\|_\infty + \left\|G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right\|\cdot\|\widehat{n}\|_2 \qquad \left(\|a\|_\infty\le\|a\|_2, \qquad \forall a\in\mathbb{C}^N\right) \\
&\le \sqrt{K}\left\|\left(G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right)G_{\rho_1}\right\|\cdot\|\alpha\|_\infty + \left\|G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right\|\cdot\|\widehat{n}\|_2 \\
&\le \sqrt{K}\left\|G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right\|\|G_{\rho_1}\|\cdot\|\alpha\|_\infty + \left\|G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right\|\cdot\|\widehat{n}\|_2 \\
&= O(1)\cdot\left\|G_{\rho_2}^\dagger-G_{\rho_1}^\dagger\right\|\cdot\left(\sqrt{K}\|\alpha\|_\infty+\|\widehat{n}\|_2\right) \qquad (\text{Lemma 15}) \\
&= O(1)\cdot\left(\|I_K-M_{\rho_1,\rho_2}\|^{\frac{1}{2}}+e^{-Cc}\right)\cdot\left(\sqrt{K}\|\alpha\|_\infty+\|\widehat{n}\|_2\right). \qquad (\text{Lemma 15}).
\end{aligned}
$$

This completes the proof of Lemma 18. □

B Hessian of $F(\cdot)$ is Positive Definite on $\mathbb{B}(\tau^0,\sigma_1)$

Throughout, assume that Proposition 2 is in force, so that $\tau^0\in\mathbb{I}^{\widetilde{K}}$ satisfies both $\widetilde{K}=K$ and $d(\tau^0,\tau)\le\sigma_1$. In this section, with $F(\cdot)$ defined as in (22), we will establish that $\frac{\partial^2 F}{\partial\rho^2}(\cdot)$ is asymptotically positive definite in the small neighborhood of τ, namely $\mathbb{B}(\tau^0,\sigma_1)$ (see (24)). (This will prove necessary for the projected Newton algorithm to converge to a local minimizer of $F(\cdot)$.) To do so, we first show that $\frac{\partial^2 F}{\partial\rho^2}(\tau)\succ 0$ asymptotically, and next control the variation of the Hessian under small changes of its argument.

We assume that the entries of $\tau\in\mathbb{I}^K$ are distinct. Then, for $\rho\in\mathbb{B}\left(\tau^0,\sigma_1\right)$, the entries of ρ too are distinct asymptotically (i.e., for large enough N). Moreover, $\rho[i]\ne\tau[j]$ for $i\ne j$ (asymptotically). Therefore, we are in position to apply the technical lemmas in the Toolbox (Appendix A).

B.1 Establishing $\frac{\partial^2 F}{\partial\rho^2}(\tau)\succ 0$

From (13), recall that $\widehat{z}_{\sigma_2}=G_\tau\alpha+\widehat{n}_{\sigma_2}\in\mathbb{C}^N$ contains the Fourier coefficients of the (possibly noisy) measurement signal. Recall also the orthogonal projection onto $\mathrm{span}(G_\tau)$, namely $\mathcal{P}_\tau\in\mathbb{C}^{N\times N}$.

Then, clearly,

$$(I_N - \mathcal{P}_\tau)\,\widehat{z}_{\sigma_2} = (I_N - \mathcal{P}_\tau)\,G_\tau\alpha + (I_N - \mathcal{P}_\tau)\,\widehat{n}_{\sigma_2} = (I_N - \mathcal{P}_\tau)\,\widehat{n}_{\sigma_2}.$$

With this in mind and using (32), we rewrite the expression for Hessian at τ as

$$\begin{aligned}\frac{\partial^2 F}{\partial \rho^2}(\tau) = {} & 2\cdot \mathrm{diag}\left(\beta_\tau\right)\cdot G_\tau^* L L^* G_\tau \cdot \mathrm{diag}\left(\beta_\tau\right) \qquad \left(L^2 = -LL^*, \quad \beta_\tau = \widetilde{\beta}(\tau) = G_\tau^\dagger \cdot \widehat{z}_{\sigma_2}\right)\\ & - 2\cdot \mathrm{diag}\left(\beta_\tau\right)\cdot \mathrm{diag}\left(G_\tau^* L^2\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\\ & - 2\left[\mathrm{diag}\left(\beta_\tau\right) G_\tau^* L G_\tau - \mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\right]\\ & \cdot \left(G_\tau^* G_\tau\right)^{-1}\cdot\left[G_\tau^* L^* G_\tau\cdot \mathrm{diag}(\beta_\tau) - \mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\right].\end{aligned}$$

After rearranging the expression above, we find that

$$\frac{\partial^2 F}{\partial \rho^2}(\tau) = \mathrm{signal}_\tau + \mathrm{noise}_\tau, \tag{70}$$

$$\mathrm{signal}_\tau := 2\cdot \mathrm{diag}\left(\beta_\tau\right)\cdot G_\tau^* L\left(I_N - \mathcal{P}_\tau\right) L^* G_\tau\cdot \mathrm{diag}\left(\beta_\tau\right), \tag{71}$$

$$\begin{aligned}\mathrm{noise}_\tau := {} & -2\cdot \mathrm{diag}\left(\beta_\tau\right)\cdot \mathrm{diag}\left(G_\tau^* L^2\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\\ & + 2\cdot \mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\cdot G_\tau^\dagger L^* G_\tau\cdot \mathrm{diag}\left(\beta_\tau\right)\\ & + 2\cdot \mathrm{diag}\left(\beta_\tau\right)\cdot\left(G_\tau^\dagger L^* G_\tau\right)^*\cdot \mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\\ & - 2\cdot \mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\cdot\left(G_\tau^* G_\tau\right)^{-1}\cdot \mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right).\end{aligned} \tag{72}$$

As detailed presently, the signal term above is "strongly" positive definite because $L^* G_\tau$ (associated with the translated copies of $g'_{\sigma_2,N}(\cdot)$, the derivative of our kernel) is nearly orthogonal to $\mathrm{span}(G_\tau)$. Therefore, as long as the noise term is negligible, we have $\frac{\partial^2 F}{\partial \rho^2}(\tau) \succ 0$. Let us consider the details now.

We first control all four terms in noise_τ. For the first term in (72), it holds asymptotically that

$$\begin{aligned}& \left\|\mathrm{diag}\left(\beta_\tau\right)\cdot \mathrm{diag}\left(G_\tau^* L^2\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\right\|\\ & \le \left\|\beta_\tau\right\|_\infty\cdot\left\|G_\tau^* L^2\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right\|_2 \qquad (\|a\|_\infty \le \|a\|_2,\ \forall a)\\ & \le \left\|\beta_\tau\right\|_\infty\cdot\left\|G_\tau\right\|\left\|L\right\|^2\left\|I_N - \mathcal{P}_\tau\right\|\cdot\left\|\widehat{n}_{\sigma_2}\right\|_2\\ & = \left\|\beta_\tau\right\|_\infty\cdot O(N^2)\cdot\left\|\widehat{n}_{\sigma_2}\right\|_2, \qquad (\text{Lemma 15, } \|L\| \le 2\pi N)\end{aligned}$$

as $c, N \to \infty$, $c = \Theta(\log N)$ (with a large enough lower bound). Similarly, for the second term in (72), it is true asymptotically that

$$\begin{aligned}& \left\|\mathrm{diag}\left(G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right)\cdot G_\tau^\dagger L^* G_\tau\cdot \mathrm{diag}\left(\beta_\tau\right)\right\|\\ & \le \left\|G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right\|_2\cdot\left\|G_\tau^\dagger\right\|\left\|L\right\|\left\|G_\tau\right\|\cdot\left\|\beta_\tau\right\|_\infty \qquad (\|a\|_\infty \le \|a\|_2,\ \forall a)\\ & \le \left\|G_\tau^* L\left(I_N - \mathcal{P}_\tau\right)\widehat{n}_{\sigma_2}\right\|_2\cdot O(N)\cdot\left\|\beta_\tau\right\|_\infty \qquad (\text{Lemma 15, } \|L\| \le 2\pi N)\\ & \le \left\|G_\tau\right\|\left\|L\right\|\left\|I_N - \mathcal{P}_\tau\right\|\cdot\left\|\widehat{n}_{\sigma_2}\right\|_2\cdot O(N)\cdot\left\|\beta_\tau\right\|_\infty\\ & \le O(N)\cdot\left\|\widehat{n}_{\sigma_2}\right\|_2\cdot O(N)\cdot\left\|\beta_\tau\right\|_\infty \qquad (\text{Lemma 15, } \|L\| \le 2\pi N)\\ & = O(N^2)\cdot\left\|\beta_\tau\right\|_\infty\left\|\widehat{n}_{\sigma_2}\right\|_2.\end{aligned}$$

An identical bound holds the third noise term. As for the last term in (72), we asymptotically have that

$$\left\|\operatorname{diag}\left(G_\tau^* L\left(I_N-\mathcal{P}_\tau\right) \widehat{n}_{\sigma_2}\right)\left(G_\tau^* G_\tau\right)^{-1} \operatorname{diag}\left(G_\tau^* L\left(I_N-\mathcal{P}_\tau\right) \widehat{n}_{\sigma_2}\right)\right\|=O(N^2) \cdot\left\|\widehat{n}_{\sigma_2}\right\|_2^2,$$

where we invoked Lemma 15 again. Overall, using the triangle inequality, we obtain that

$$\left\|\text{noise}_\tau\right\|=O(N^2) \cdot\left(\left\|\beta_\tau\right\|_{\infty}\left\|\widehat{n}_{\sigma_2}\right\|_2+\left\|\widehat{n}_{\sigma_2}\right\|_2^2\right). \qquad \text{(see (72))} \tag{73}$$

To eliminate β_τ from the expression above, we apply Lemma 18 (with $\rho_1=\rho_2=\tau$) to obtain that

$$\begin{aligned}\left\|\beta_\tau\right\|_{\infty} & \leq\|\alpha\|_{\infty}+\left\|\beta_\tau-\alpha\right\|_{\infty} \\ & \leq 2\|\alpha\|_{\infty}+O(1)\left\|\widehat{n}_{\sigma_2}\right\|_2 . \qquad \left(M_{\tau, \tau}=I_K\right)\end{aligned} \tag{74}$$

Therefore,

$$\begin{aligned}\left\|\text{noise}_\tau\right\| & =O(N^2) \cdot\left(\left\|\beta_\tau\right\|_{\infty}\left\|\widehat{n}_{\sigma_2}\right\|_2+\left\|\widehat{n}_{\sigma_2}\right\|_2^2\right), \qquad \text{(see (73))} \\ & =O(N^2)\left(\left(\|\alpha\|_{\infty}+\left\|\widehat{n}_{\sigma_2}\right\|_2\right)\left\|\widehat{n}_{\sigma_2}\right\|_2+\left\|\widehat{n}_{\sigma_2}\right\|_2^2\right) \qquad \text{(see (74))} \\ & =O(N^2)\left(\|\alpha\|_{\infty}\left\|\widehat{n}_{\sigma_2}\right\|_2+\left\|\widehat{n}_{\sigma_2}\right\|_2^2\right),\end{aligned}$$

and, therefore,

$$\text{noise}_\tau \preccurlyeq O(N^2)\left(\|\alpha\|_{\infty}\left\|\widehat{n}_{\sigma_2}\right\|_2+\left\|\widehat{n}_{\sigma_2}\right\|_2^2\right) \cdot I_K, \tag{75}$$

both valid asymptotically. Next, we establish that the signal term in (70) is a positive definite matrix. For arbitrary $v \in \mathbb{R}^K$, it holds asymptotically that

$$\begin{aligned}
& v^*\left(G_\tau^* L\left(I_N-\mathcal{P}_\tau\right) L^* G_\tau\right) v \\
& =v^*\left(G_\tau^* L L^* G_\tau\right) v-v^*\left(G_\tau^* L \mathcal{P}_\tau L^* G_\tau\right) v \\
& =\left\|L^* G_\tau v\right\|_2^2-\left\|\mathcal{P}_\tau L^* G_\tau v\right\|_2^2 \qquad \left(\mathcal{P}_\tau^2=\mathcal{P}_\tau\right) \\
& =\left\|L G_\tau v\right\|_2^2-\left\|\left(G_\tau^{\dagger}\right)^* G_\tau^* L^* G_\tau v\right\|_2^2 \qquad \left(L^*=-L, \quad \mathcal{P}_\tau=\left(G_\tau^{\dagger}\right)^* G_\tau^*\right) \\
& \geq\left\|L G_\tau v\right\|_2^2-\left\|G_\tau^{\dagger}\right\|^2 \cdot\left\|G_\tau^* L^* G_\tau v\right\|_2^2 \\
& =\left\|L G_\tau v\right\|_2^2-O(1) \cdot\left\|G_\tau^* L^* G_\tau\right\|^2 \cdot\|v\|_2^2 \qquad \text{(see (55))} \\
& \geq\left(\left\|g_{\sigma, N}^{\prime}(\cdot)\right\|_2^2-O(e^{-C c_2})\right) \cdot\|v\|_2^2-O(1) \cdot O(e^{-C c_2}) \cdot\|v\|_2^2 \qquad \text{(Lemmas 15 and 16)} \\
& =\Omega(N^2) \cdot\|v\|_2^2, \qquad \text{(Criterion 5)}
\end{aligned}$$

as $c_2, N \rightarrow \infty$ and $c_2=\Theta(\log N)$ (with large enough lower bound). Since the choice of v was arbitrary, we conclude that

$$G_\tau^* L\left(I_N-\mathcal{P}_\tau\right) L^* G_\tau \succcurlyeq \Omega(N^2) \cdot I_K,$$

asymptotically. From (71), it follows that

$$\text{signal}_\tau=2 \cdot \operatorname{diag}\left(\beta_\tau\right) \cdot G_\tau^* L\left(I_N-\mathcal{P}_\tau\right) L^* G_\tau \cdot \operatorname{diag}\left(\beta_\tau\right) \succcurlyeq \Omega(N^2) \cdot \operatorname{diag}\left(\beta_\tau\right)^2 . \tag{76}$$

We remove β_τ from the right hand side above by invoking Lemma 18: Note that

$$\begin{aligned}\operatorname{diag}(\beta_\tau) &= \operatorname{diag}(\alpha) - \operatorname{diag}(\alpha - \beta_\tau)\\ &\succcurlyeq \operatorname{diag}(\alpha) - \|\alpha - \beta_\tau\|_\infty \cdot I_K,\\ &\succcurlyeq \frac{1}{2}\operatorname{diag}(\alpha) - O(1)\cdot \|\widehat{n}_{\sigma_2}\|_2 \cdot I_K,\end{aligned}$$

$$\operatorname{diag}(\beta_\tau)^2 \succcurlyeq \frac{1}{8}\operatorname{diag}(\alpha)^2 - O(1)\cdot \|\widehat{n}_{\sigma_2}\|_2^2 \cdot I_K, \qquad \left((a-b)^2 \geq \frac{a^2}{2} - b^2, \quad \forall a,b\in\mathbb{R}\right),$$

asymptotically. Therefore, revisiting (76), we can write that

$$\begin{aligned}\text{signal}_\tau &\succcurlyeq \Omega(N^2)\cdot \operatorname{diag}(\beta_\tau)^2\\ &\succcurlyeq \Omega(N^2)\cdot\left(\operatorname{diag}(\alpha)^2 - O(1)\cdot\|\widehat{n}_{\sigma_2}\|_2^2\cdot I_K\right). \end{aligned}\tag{77}$$

Suppose that

$$\|\widehat{n}_{\sigma_2}\|_2 \leq O(1)\cdot\|\alpha\|_\infty, \tag{78}$$

with a small enough constant. Then, combining (77) with (75) yields

$$\begin{aligned}&\frac{\partial^2 F}{\partial \rho^2}(\tau)\\ &= \text{signal}_\tau + \text{noise}_\tau \qquad \text{(see (70))}\\ &\succcurlyeq \Omega(N^2)\cdot\left(\operatorname{diag}(\alpha)^2 - O(1)\cdot\|\widehat{n}_{\sigma_2}\|_2^2\cdot I_K\right) - O(N^2)\cdot\left(\|\alpha\|_\infty\|\widehat{n}_{\sigma_2}\|_2 + \|\widehat{n}_{\sigma_2}\|_2^2\right)\cdot I_K\\ &\succcurlyeq \Omega(N^2)\cdot\operatorname{diag}(\alpha)^2 - O(N^2)\cdot\|\alpha\|_\infty\|\widehat{n}_{\sigma_2}\|_2\cdot I_K \qquad \text{(see (78))}\\ &= \Omega(N^2)\cdot\|\alpha\|_\infty^2\cdot\frac{\operatorname{diag}(\alpha)^2}{\|\alpha\|_\infty^2} - O(N^2)\cdot\|\alpha\|_\infty\|\widehat{n}_{\sigma_2}\|_2\cdot I_K\\ &\succcurlyeq \Omega(N^2)\cdot\|\alpha\|_\infty^2\cdot\frac{\min_i|\alpha[i]|^2}{\max_i|\alpha[i]|^2}\cdot I_K - O(N^2)\cdot\|\alpha\|_\infty\|\widehat{n}_{\sigma_2}\|_2\cdot I_K\\ &= \Omega(N^2)\cdot\|\alpha\|_\infty^2\cdot\left(\operatorname{dyn}(x_{\tau,\alpha})^{-2} - O(1)\cdot\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right)\cdot I_K, \qquad \text{(see (10))}\end{aligned}\tag{79}$$

asymptotically. Therefore, as long as

$$\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty} = O(1)\cdot\operatorname{dyn}(x_{\tau,\alpha})^{-2} \leq 1,$$

with a small enough constant, $\frac{\partial^2 F}{\partial\rho^2}(\tau)\succ 0$ asymptotically (as we hoped to establish).

B.2 Establishing $\frac{\partial^2 F}{\partial\rho^2}(\rho)\succ 0$ When ρ is Close to τ

It should be clear that, by continuity,

$$\frac{\partial^2 F}{\partial\rho^2}(\tau)\succ 0 \Longrightarrow \frac{\partial^2 F}{\partial\rho^2}(\rho)\succ 0,$$

when $\rho \in \mathbb{I}^K$ is sufficiently close to τ. In this section, we precisely calculate the neighborhood of τ in $\mathbb{I}^K$ over which the Hessian of $F(\cdot)$ is positive definite. To that end, for $\rho \in \mathbb{B}(\tau, 2\sigma_1)$, we write that

$$\frac{\partial^2 F}{\partial \rho^2}(\rho) = \frac{\partial^2 F}{\partial \rho^2}(\tau) + \left(\frac{\partial^2 F}{\partial \rho^2}(\rho) - \frac{\partial^2 F}{\partial \rho^2}(\tau) \right),$$

and, to control the variation, note that

$$\begin{aligned} \left\| \frac{\partial^2 F}{\partial \rho^2}(\rho) - \frac{\partial^2 F}{\partial \rho^2}(\tau) \right\| &= \left\| \left(\text{signal}_\rho + \text{noise}_\rho\right) - \left(\text{signal}_\tau + \text{noise}_\tau\right) \right\| \qquad \text{(see (70))} \\ &\le \left\| \text{signal}_\rho - \text{signal}_\tau \right\| + \left\| \text{noise}_\rho \right\| + \left\| \text{noise}_\tau \right\|. \end{aligned} \tag{80}$$

We begin by comparing the signal terms of the Hessian at ρ and τ in the asymptotic regime $c, N \to \infty$ and $c = \Theta(\log N)$ (with a large enough lower bound). Below, we repeatedly use the identity $AB - CD = (A-C)D + A(B-D)$ for conformal matrices A, B, C, D. After recalling (71), we write that

$$\begin{aligned} &\left\| \text{signal}_\rho - \text{signal}_\tau \right\| \\ &= 2 \left\| \overbrace{\operatorname{diag}(\beta_\rho) \cdot G_\rho^* L}^{A} \cdot \overbrace{(I_N - \mathcal{P}_\rho)\, L^* G_\rho \cdot \operatorname{diag}(\beta_\rho)}^{B} \right. \\ &\qquad \left. - \overbrace{\operatorname{diag}(\beta_\tau) \cdot G_\tau^* L}^{C} \cdot \overbrace{(I_N - \mathcal{P}_\tau)\, L^* G_\tau \cdot \operatorname{diag}(\beta_\tau)}^{D} \right\| \\ &\le 4 \left\| L^* G_\rho \cdot \operatorname{diag}(\beta_\rho) - L^* G_\tau \cdot \operatorname{diag}(\beta_\tau) \right\| \cdot \max\left[\left\| L^* G_\rho \cdot \operatorname{diag}(\beta_\rho) \right\|, \left\| L^* G_\tau \cdot \operatorname{diag}(\beta_\tau) \right\| \right] \\ &\le 16\pi^2 N^2 \left\| \overbrace{G_\rho}^{A} \cdot \overbrace{\operatorname{diag}(\beta_\rho)}^{B} - \overbrace{G_\tau}^{C} \cdot \overbrace{\operatorname{diag}(\beta_\tau)}^{D} \right\| \\ &\qquad \cdot \max\left[\|G_\rho\| \, \|\beta_\rho\|_\infty, \|G_\tau\| \, \|\beta_\tau\|_\infty \right], \qquad (\|L\| \le 2\pi N) \end{aligned}$$

and, consequently,

$$\begin{aligned} &\left\| \text{signal}_\rho - \text{signal}_\tau \right\| \\ &\le 16\pi^2 N^2 \left(\|G_\rho - G_\tau\| \, \|\beta_\tau\|_\infty + \|G_\rho\| \, \|\beta_\rho - \beta_\tau\|_\infty \right) \cdot \max\left[\|G_\rho\| \, \|\beta_\rho\|_\infty, \|G_\tau\| \, \|\beta_\tau\|_\infty \right] \\ &= O(N^2) \left(\left(\|I_K - M_{\rho,\tau}\|^{\frac{1}{2}} + e^{-Cc} \right) \|\beta_\tau\|_\infty + \|\beta_\rho - \beta_\tau\|_\infty \right) \\ &\qquad \cdot \max\left[\|\beta_\rho\|_\infty, \|\beta_\tau\|_\infty \right]. \qquad \text{(Lemma 15)} \end{aligned} \tag{81}$$

We further simplify the last line above as follows. Since $\rho \in \mathbb{B}(\tau, 2\sigma_1)$ and $2\sigma_1 \le h(\sigma_2, N) \le \sigma_2$ (all by hypothesis), Criterion 5 is in force and, asymptotically, we may write that

$$\begin{aligned} &\|I_K - M_{\rho,\tau}\| \\ &= \max_{i \in [1:K]} \left| 1 - \langle g_{\sigma_2,N}(t \ominus \rho[i]), g_{\sigma_2,N}(t \ominus \tau[i]) \rangle \right| \qquad \text{(see (59))} \\ &= O(1) \cdot \max_{i \in [1:K]} d\left(\rho[i], \tau[i]\right) \qquad \text{(Criterion 5)} \\ &= O(1) \cdot d(\rho, \tau). \qquad \text{(see (8))} \end{aligned} \tag{82}$$

Moreover, to remove the terms involving β_τ and β_ρ in (81), we invoke Lemma 18 to write the following asymptotic estimates:

$$\begin{aligned}\|\beta_\rho-\beta_\tau\|_\infty &= O(1)\cdot\left(\|I_K-M_{\rho,\tau}\|^{\frac{1}{2}}+e^{-Cc_2}\right)\cdot\left(\sqrt{K}\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right)\\ &= O(1)\cdot\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\cdot\left(\sqrt{K}\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right), \qquad \text{(see (82))}\end{aligned} \tag{83}$$

$$\begin{aligned}\|\beta_\rho\|_\infty &\le \|\alpha\|_\infty+\|\beta_\rho-\alpha\|_\infty\\ &= \|\alpha\|_\infty+O(1)\cdot\left(\sqrt{K}\,\|I_K-M_{\rho,\tau}\|^{\frac{1}{2}}+e^{-Cc_2}\right)\cdot\|\alpha\|_\infty+O(1)\,\|\widehat{n}_{\sigma_2}\|_2\\ &= O(1)\cdot\left(\left(1+\sqrt{K}\cdot d(\rho,\tau)^{\frac{1}{2}}\right)\cdot\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right).\end{aligned} \tag{84}$$

Using the estimates above, we revisit (81):

$$\begin{aligned}&\left\|\text{signal}_\rho-\text{signal}_\tau\right\|\\ &= O(N^2)\left(\left(\|I_K-M_{\rho,\tau}\|^{\frac{1}{2}}+e^{-Cc_2}\right)\|\beta_\tau\|_\infty+\|\beta_\rho-\beta_\tau\|_\infty\right)\cdot\max\left[\|\beta_\rho\|_\infty,\|\beta_\tau\|_\infty\right]\\ &= O(N^2)\left(\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)+\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\right)\left(\sqrt{K}\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right)\\ &\quad\cdot\left(\left(1+\sqrt{K}\cdot d(\rho,\tau)^{\frac{1}{2}}\right)\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right) \qquad \text{(see (74) , (82-84))}\\ &= O(N^2)\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\left(\sqrt{K}\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right)^2\ (\rho\in\mathbb{B}(\tau,\sigma_1)\Longrightarrow K\cdot d(\rho,\tau)=o(1))\\ &= O(N^2)\cdot\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\cdot\left(K\|\alpha\|_\infty^2+\|\widehat{n}_{\sigma_2}\|_2^2\right) \qquad \left((a+b)^2\le 2a^2+2b^2,\quad \forall a,b\in\mathbb{R}\right)\\ &= O(N^2)\cdot\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\cdot\left(K\|\alpha\|_\infty^2+\|\alpha\|_\infty\|\widehat{n}_{\sigma_2}\|_2\right). \qquad \left(\text{if } \|\widehat{n}_{\sigma_2}\|_2=O(1)\cdot\|\alpha\|_\infty\right)\end{aligned} \tag{85}$$

It remains to control noise_τ and noise_ρ in (80). In the analysis that started in (72) and led to (75), we earlier bounded $\|\text{noise}_\tau\|$. So we turn our attention to noise_ρ. Note that

$$\begin{aligned}\widehat{z}_{\sigma_2} &= G_\tau\alpha+\widehat{n}_{\sigma_2} \qquad \text{(see (13))}\\ &= G_\rho\alpha+\widehat{n}_{\sigma_2}+(G_\tau-G_\rho)\,\alpha =: G_\rho\alpha+\widehat{n}'_{\sigma_2},\end{aligned}$$

$$\begin{aligned}\beta_\rho &= G_\rho^\dagger\widehat{z}_{\sigma_2}\\ &= G_\rho^\dagger\left(G_\rho\alpha+\widehat{n}'_{\sigma_2}\right),\end{aligned}$$

$$\left(I_N-\mathcal{P}_\rho\right)\widehat{z}_{\sigma_2}=\left(I_N-\mathcal{P}_\rho\right)\widehat{n}'_{\sigma_2},$$

so that we next apply (75) but with ρ and $\widehat{n}'_{\sigma_2}=\widehat{n}_{\sigma_2}+(G_\tau-G_\rho)\alpha$ (instead of τ and $\widehat{n}_{\sigma_2}$) to obtain that

$$\begin{aligned}&\|\text{noise}_\rho\|\\ &= O(N^2)\cdot\left[\|\beta_\rho\|_\infty\left\|\widehat{n}'_{\sigma_2}\right\|_\infty+\left\|\widehat{n}'_{\sigma_2}\right\|_\infty^2\right]\\ &= O(N^2)\cdot\|\beta_\rho\|_\infty\left\|\widehat{n}'_{\sigma_2}\right\|_\infty \qquad \left(\text{if } \left\|\widehat{n}'_{\sigma_2}\right\|_\infty\le\|\beta_\rho\|_\infty\right)\\ &= O(N^2)\cdot\|\beta_\rho\|_\infty\left\|\widehat{n}_{\sigma_2}+(G_\tau-G_\rho)\,\alpha\right\|_\infty\\ &\le O(N^2)\cdot\left(\left(1+\sqrt{K}\cdot d(\rho,\tau)^{\frac{1}{2}}\right)\cdot\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right)\\ &\qquad\cdot\left(\|\widehat{n}_{\sigma_2}\|_2+\|G_\tau-G_\rho\|_{\infty\to\infty}\|\alpha\|_\infty\right), \qquad \text{(see (84))}\end{aligned}$$

and, consequently,

$$
\begin{aligned}
&\|\text{noise}_\rho\| \\
&= O(N^2)\cdot\left(\left(1+\sqrt{K}\cdot d(\rho,\tau)^{\frac{1}{2}}\right)\cdot\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right) \\
&\quad\cdot\left(\|\widehat{n}_{\sigma_2}\|_2+\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\|\alpha\|_\infty\right) \text{(see (62) and (82))} \\
&= O(N^2)\cdot\left(\left(\sqrt{K}\cdot d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\|\alpha\|_\infty+\|\widehat{n}_{\sigma_2}\|_2\right)^2 \quad \left(d(\rho,\tau)\le\frac{1}{2},\ e^{-Cc_2}=o(1)\right) \\
&= O(N^2)\cdot\left(\left(K\cdot d(\rho,\tau)+e^{-Cc_2}\right)\|\alpha\|_\infty^2+\|\widehat{n}_{\sigma_2}\|_2^2\right) \qquad \left((a+b)^2\le 2a^2+2b^2,\quad\forall a,b\in\mathbb{R}\right) \\
&= O(N^2)\cdot\left(\left(K\cdot d(\rho,\tau)+e^{-Cc_2}\right)\|\alpha\|_\infty^2+\|\alpha\|_\infty\|\widehat{n}_{\sigma_2}\|_2\right). \qquad (\text{if } \|\widehat{n}_{\sigma_2}\|_2\le\|\alpha\|_\infty)
\end{aligned} \tag{86}
$$

Using Lemmas 15 and 18, it is not difficult to verify that both conditions imposed while deriving (86) hold if

$$
\|\widehat{n}_{\sigma_2}\|_2 = O(1)\cdot\frac{\|\alpha\|_\infty}{1-\sqrt{K}\cdot d(\rho,\tau)^{\frac{1}{2}}},
$$

with a small enough constant. Since $K\cdot d(\rho,\tau)=o(1)$ for any $\rho\in\mathbb{B}(\tau,2\sigma_1)$, the condition above is met asymptotically when $\|\widehat{n}_{\sigma_2}\|_2=O(1)\|\alpha\|_\infty$ (with a small enough constant). In light of (80), we can combine the estimates above to obtain that

$$
\begin{aligned}
&\left\|\frac{\partial^2 F}{\partial\rho^2}(\rho)-\frac{\partial^2 F}{\partial\rho^2}(\tau)\right\| \\
&\le \left\|\text{signal}_\rho-\text{signal}_\tau\right\|+\|\text{noise}_\rho\|+\|\text{noise}_\tau\| \\
&= O(N^2)\cdot\|\alpha\|_\infty^2\cdot\left(d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}\right)\cdot\left(K+\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right) \\
&\quad+O(N^2)\cdot\|\alpha\|_\infty^2\cdot\left(K\cdot d(\rho,\tau)+e^{-Cc_2}+\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right) \qquad \text{(see (75), (85), and (86))} \\
&= O(N^2)\cdot\|\alpha\|_\infty^2 \\
&\quad\cdot\left[K\cdot d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}+\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right]. \qquad \left(c=\Theta(\log N)\Rightarrow Ke^{-Cc_2}=O\left(e^{-Cc_2}\right)\right)
\end{aligned} \tag{87}
$$

The lower bound in $c=\Theta(\log N)$ must be sufficiently large for the last line above to hold. It immediately follows that

$$
\frac{\partial^2 F}{\partial\rho^2}(\rho)\succcurlyeq\frac{\partial^2 F}{\partial\rho^2}(\tau)-\left\|\frac{\partial^2 F}{\partial\rho^2}(\rho)-\frac{\partial^2 F}{\partial\rho^2}(\tau)\right\|\cdot I_K,
$$

and, consequently,

$$
\begin{aligned}
&\frac{\partial^2 F}{\partial\rho^2}(\rho) \\
&\succcurlyeq \Omega(N^2)\cdot\|\alpha\|_\infty^2\cdot\left[\operatorname{dyn}\left(x_{\tau,\alpha}\right)^{-2}-O(1)\cdot\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right]\cdot I_K \\
&\qquad -O(N^2)\cdot\|\alpha\|_\infty^2\cdot\left[K\cdot d(\rho,\tau)^{\frac{1}{2}}+e^{-Cc_2}+\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right]\cdot I_K \qquad \text{(see (79) and (87))} \\
&= \Omega(N^2)\cdot\|\alpha\|_\infty^2\cdot\left[\operatorname{dyn}\left(x_{\tau,\alpha}\right)^{-2}-O(K)\cdot d(\rho,\tau)^{\frac{1}{2}}-O\left(e^{-Cc_2}\right)-O(1)\cdot\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\right]\cdot I_K,
\end{aligned}
$$

asymptotically. The component $O(K)\cdot d(\rho,\tau)^{\frac{1}{2}}+O(e^{-Cc_2})$ in the last line above is asymptotically negligible. Indeed, $d(\rho,\tau)=o(1)$ when $\rho\in\mathbb{B}(\tau^0,\sigma_1)$ (with $\sigma_1=\frac{c_1}{N}$), and also $e^{-Cc_2}=o(1)$. Therefore, as long as

$$\frac{\|\widehat{n}_{\sigma_2}\|_2}{\|\alpha\|_\infty}\leq\frac{\|n(\cdot)\|_{L_2}}{\|\alpha\|_\infty}=\frac{O(1)}{\mathrm{dyn}\left(x_{\tau,\alpha}\right)^2}, \tag{88}$$

and with a small enough constant, $\frac{\partial^2 F}{\partial\rho^2}(\rho)\succ 0$ holds asymptotically for every $\rho\in\mathbb{B}\left(\tau,2\sigma_1\right)$. In particular, because

$$\rho\in\mathbb{B}\left(\tau^0,\sigma_1\right)\Longrightarrow d\left(\rho,\tau\right)\leq d\left(\rho,\tau^0\right)+d\left(\tau^0,\tau\right)\leq 2\sigma_1,$$

we asymptotically have that $\frac{\partial^2 F}{\partial\rho^2}(\rho)\succ 0$ for every $\rho\in\mathbb{B}\left(\tau^0,\sigma_1\right)$. The inequality in (88) is established next:

$$\begin{aligned}\|\widehat{n}_{\sigma_2}\|_2&=\left\|\left(g_{\sigma_2,N}\circledast n\right)(\cdot)\right\|_{L_2}\qquad\text{(Parseval's identity and (11))}\\&\leq\left\|g_{\sigma_2,N}(\cdot)\right\|_{L_2}\|n(\cdot)\|_{L_2}\\&=\|n(\cdot)\|_{L_2}.\qquad\text{(see Criterion 4)}\end{aligned} \tag{89}$$

C Proof of Lemma 12

Recall that the entries of $\tau\in\mathbb{I}^K$ are distinct. Because $\rho\in\mathbb{B}\left(\tau^0,\sigma_1\right)$ (by Proposition 2), the entries of ρ too are distinct asymptotically (i.e., for large enough N). Moreover, $\rho[i]\neq\tau[j]$ for $i\neq j$. Therefore, we are in position to apply the technical lemmas in the Toolbox Section.

With $u\in\mathbb{I}^K$ to be set later, let $U=\mathrm{diag}(u)\in\mathbb{R}^{K\times K}$ be the diagonal matrix formed by the vector u. Using the expression for the gradient of $F(\cdot)$ from the accompanying document [11], we can write that

$$\begin{aligned}&\left\langle\frac{\partial F}{\partial\rho}(\rho),-u\right\rangle\\&=\left\langle-\frac{\partial F}{\partial\rho}(\rho),u\right\rangle\\&=\left\langle\mathrm{diag}(\beta_\rho)\cdot G_\rho^*L\left(\widehat{z}_{\sigma_2}-G_\rho\beta_\rho\right),u\right\rangle,\qquad\left(\beta_\rho=\widetilde{\beta}(\rho)=G_\rho^\dagger\widehat{z}_{\sigma_2}\right)\end{aligned}$$

and, consequently,

$$\begin{aligned}&\left\langle\frac{\partial F}{\partial\rho}(\rho),-u\right\rangle\\&=\left\langle\mathrm{diag}(\beta_\rho)\cdot G_\rho^*L\left(G_\tau\alpha+\widehat{n}_{\sigma_2}-G_\rho\beta_\rho\right),u\right\rangle\qquad\text{(see (13))}\\&=\langle LG_\tau\alpha+L\widehat{n}_{\sigma_2}-LG_\rho\beta_\rho,G_\rho U\beta_\rho\rangle\\&=\left\langle LG_\tau\alpha+L\widehat{n}_{\sigma_2}-LG_\rho G_\rho^\dagger\left(G_\tau\alpha+\widehat{n}_{\sigma_2}\right),G_\rho UG_\rho^\dagger\left(G_\tau\alpha+\widehat{n}_{\sigma_2}\right)\right\rangle\\&=\langle L\left(I_N-\mathcal{P}_\rho\right)G_\tau\alpha+L\left(I_N-\mathcal{P}_\rho\right)\widehat{n}_{\sigma_2},\mathcal{P}_{\rho,U}G_\tau\alpha+\mathcal{P}_{\rho,U}\widehat{n}_{\sigma_2}\rangle\;\left(\mathcal{P}_{\rho,U}:=G_\rho UG_\rho^\dagger,\quad\mathcal{P}_\rho=\mathcal{P}_{\rho,I}\right)\\&=\underbrace{\langle LG_\tau\alpha,\mathcal{P}_{\rho,U}G_\tau\alpha\rangle-\langle L\mathcal{P}_\rho G_\tau\alpha,\mathcal{P}_{\rho,U}G_\tau\alpha\rangle}_{\text{signal terms}}\\&+\underbrace{\langle L\left(I_N-\mathcal{P}_\rho\right)\widehat{n}_{\sigma_2},\mathcal{P}_{\rho,U}G_\tau\alpha\rangle+\langle L\left(I_N-\mathcal{P}_\rho\right)G_\tau\alpha,\mathcal{P}_{\rho,U}\widehat{n}_{\sigma_2}\rangle+\langle L\left(I_N-\mathcal{P}_\rho\right)\widehat{n}_2,\mathcal{P}_{\rho,U}\widehat{n}_{\sigma_2}\rangle}_{\text{noise terms}}.\end{aligned} \tag{90}$$

In order to find a lower bound for the inner product $\langle -\frac{\partial F}{\partial \rho}(\rho), u\rangle$, we will study each of the five terms in the last identity in (90). The first term there can be approximated with a simpler quantity as follows. Asymptotically, we have that

$$\begin{aligned}
&\left|\langle LG_\tau\alpha, \mathcal{P}_{\rho,U}G_\tau\alpha\rangle - \left\langle M^d_{\rho,\tau}\alpha, M_{\rho,\tau}U\alpha\right\rangle\right| \qquad \text{(see (59) and (61))}\\
&\le |\langle LG_\tau\alpha, \mathcal{P}_{\rho,U}G_\tau\alpha\rangle - \langle LG_\tau\alpha, G_\rho U M_{\rho,\tau}\alpha\rangle| + \left|\langle LG_\tau\alpha, G_\rho U M_{\rho,\tau}\alpha\rangle - \left\langle M^d_{\rho,\tau}\alpha, M_{\rho,\tau}U\alpha\right\rangle\right|\\
&\le \|LG_\tau\alpha\|_2 \cdot \|\mathcal{P}_{\rho,U}G_\tau\alpha - G_\rho U M_{\rho,\tau}\alpha\|_2 + \left\|G_\rho^* L G_\tau - M^d_{\rho,\tau}\right\| \|u\|_\infty \|M_{\rho,\tau}\|_\infty \|\alpha\|_2^2\\
&= \|LG_\tau\alpha\|_2 \cdot \|\mathcal{P}_{\rho,U}G_\tau\alpha - G_\rho U M_{\rho,\tau}\alpha\|_2\\
&\qquad + O(e^{-Cc_2})\|u\|_\infty \|\alpha\|_2^2 \qquad \text{(see (60) and text below)}\\
&= \|LG_\tau\alpha\|_2\, O(e^{-Cc_2})\|u\|_\infty \|G_\tau\alpha\|_2 + O(e^{-Cc_2})\|u\|_\infty\|\alpha\|_2^2 \qquad \text{(Lemma 17)}\\
&= \|L\| \cdot O(e^{-Cc_2})\|u\|_\infty \|\alpha\|_2^2 + O(e^{-Cc_2})\|u\|_\infty\|\alpha\|_2^2 \qquad \text{(see (53))}\\
&= O(e^{-Cc_2})\|u\|_\infty \|\alpha\|_2^2. \qquad (\|L\| \le 2\pi N,\ c_2 = \Theta(\log N))
\end{aligned} \tag{91}$$

In the fourth line above, $\|M_{\rho,\tau}\|_\infty$ is absorbed as a constant on account of the asymptotic bound

$$\|M_{\rho,\tau}\|_\infty = \|M_{\rho,\tau}\| \le \left\|G_\rho^* G_\tau\right\| + O(e^{-Cc_2}) \le \|G_\rho\|\,\|G_\tau\| + O(e^{-Cc_2}) \le 2,$$

which holds because $M_{\rho,\tau}$ is diagonal and by Lemma 15 (see (53) and (58)). In the last line of (91), the lower bound in $c_2 = \Theta(\log N)$ must be sufficiently large. Next, we can asymptotically upper-bound the second term in the last identity in (90) as follows:

$$\begin{aligned}
|\langle L\mathcal{P}_\rho G_\tau\alpha, \mathcal{P}_{\rho,U}G_\tau\alpha\rangle| &\le \|\mathcal{P}_{\rho,U} L \mathcal{P}_\rho\| \|G_\tau\alpha\|_2^2\\
&= \left\|\left[\left(G_\rho^\dagger\right)^* U G_\rho^*\right] L \left[G_\rho G_\rho^\dagger\right]\right\| \|G_\tau\alpha\|_2^2 \qquad \left(\mathcal{P}_{\rho,U} = \mathcal{P}_{\rho,U}^* = G_\rho U G_\rho^\dagger\right)\\
&\le \left\|G_\rho^\dagger\right\| \cdot \|u\|_\infty \cdot \left\|G_\rho^* L G_\rho\right\| \cdot \left\|G_\rho^\dagger\right\| \cdot \|G_\tau\alpha\|_2^2\\
&= O(e^{-Cc_2})\|u\|_\infty\|\alpha\|_2^2. \qquad \text{(Lemmas 15 and 16)}
\end{aligned} \tag{92}$$

These lemmas are applicable because the entries of τ and ρ are each distinct. Similarly, we asymptotically upper-bound the third term on the last identity in (90) as follows:

$$\begin{aligned}
&|\langle L\left(I_N - \mathcal{P}_\rho\right)\widehat{n}_{\sigma_2}, \mathcal{P}_{\rho,U}G_\tau\alpha\rangle|\\
&\le \|L\|\,\|I_N - \mathcal{P}_\rho\|\,\|\widehat{n}_{\sigma_2}\|_2\,\|\mathcal{P}_{\rho,U}\|\,\|G_\tau\alpha\|_2\\
&= O(N)\,\|\widehat{n}_{\sigma_2}\|_2\,\|u\|_\infty\,\|\alpha\|_2. \qquad (\|L\| \le 2\pi N,\ \|I_N - \mathcal{P}_\rho\| \le 1,\ \text{Lemmas 16 and 17})
\end{aligned} \tag{93}$$

Next, consider the fourth term in the last identity in (90). Asymptotically, it holds that

$$\begin{aligned}
|\langle L\left(I_N - \mathcal{P}_\rho\right)G_\tau\alpha, \mathcal{P}_{\rho,U}\widehat{n}_{\sigma_2}\rangle| &\le \|L\|\,\|I_N - \mathcal{P}_\rho\|\,\|G_\tau\alpha\|_2\,\|\mathcal{P}_{\rho,U}\|\,\|\widehat{n}_{\sigma_2}\|_2\\
&= O(N)\,\|\alpha\|_2\,\|u\|_\infty\,\|\widehat{n}_{\sigma_2}\|_2,
\end{aligned} \tag{94}$$

with a similar argument. Finally, consider the fifth term on the last line of (90):

$$\begin{aligned}
|\langle L\left(I_N - \mathcal{P}_\rho\right)\widehat{n}_{\sigma_2}, \mathcal{P}_{\rho,U}\widehat{n}_{\sigma_2}\rangle| &\le \|L\|\,\|I_N - \mathcal{P}_\rho\|\,\|\mathcal{P}_{\rho,U}\|_2\,\|\widehat{n}_{\sigma_2}\|_2^2\\
&= O(N)\,\|u\|_\infty\,\|\widehat{n}_{\sigma_2}\|_2^2.
\end{aligned} \tag{95}$$

We now use (91-95) to find a lower bound for the inner product in (90):

$$\left\langle -\frac{\partial F}{\partial \rho}(\rho), u \right\rangle \geq \left\langle M^d_{\rho,\tau}\alpha, M_{\rho,\tau}U\alpha \right\rangle - O(e^{-Cc_2})\|u\|_\infty \|\alpha\|_2^2 \\ - O(N)\|u\|_\infty \left\|\widehat{n}_{\sigma_2}\right\|_2 \|\alpha\|_2 - O(N) \|u\|_\infty \left\|\widehat{n}_{\sigma_2}\right\|_2^2 . \tag{96}$$

Let us simplify the lower bound above. To that end, observe that

$$\left\langle M^d_{\rho,\tau}\alpha, M_{\rho,\tau}U\alpha \right\rangle = \sum_{i=1}^{K} |\alpha[i]|^2 \cdot u[i] \cdot M_{\rho,\tau}[i,i] \cdot M^d_{\rho,\tau}[i,i], \tag{97}$$

which owes itself to the fact that U, $M_{\rho,\tau}$, and $M^d_{\rho,\tau}$ are all diagonal matrices. First, by design,

$$\rho, \tau \in \mathbb{B}(\tau^0, \sigma_1) \Rightarrow d(\rho,\tau) \leq 2\sigma_1 \leq h(\sigma_2, N).$$

Then, on the account of Criterion 5, we asymptotically have that

$$M_{\rho,\tau}[i,i] = \langle g_{\sigma_2,N}(t \ominus \rho[i]), g_{\sigma_2,N}(t \ominus \tau[i]) \rangle = \Omega(1), \tag{98}$$

$$\begin{aligned} \left| M^d_{\rho,\tau}[i,i] \right| &= \left| \langle g_{\sigma_2,N}(t \ominus \rho[i]), g'_{\sigma_2,N}(t \ominus \tau[i]) \rangle \right| \\ &= \text{sign}\left(\rho[i] \ominus \tau[i] - \frac{1}{2} \right) \cdot \langle g_{\sigma_2,N}(t \ominus \rho[i]), g'_{\sigma_2,N}(t \ominus \tau[i]) \rangle \\ &= \Omega(N^2) \cdot d(\rho[i], \tau[i]). \end{aligned} \tag{99}$$

Second, we choose

$$u = \text{sign}\left((\rho \ominus \tau) - \frac{1}{2} \right).$$

With this choice of u, it asymptotically holds that

$$\begin{aligned} &\left\langle M^d_{\rho,\tau}\alpha, M_{\rho,\tau}U\alpha \right\rangle \\ &= \sum_{i=1}^{K} |\alpha[i]|^2 \cdot u[i] \cdot M_{\rho,\tau}[i,i] \cdot M^d_{\rho,\tau}[i,i] \qquad \text{(see (97))} \\ &= \sum_{i=1}^{K} |\alpha[i]|^2 \cdot \left| M^d_{\rho,\tau}[i,i] \right| \cdot M_{\rho,\tau}[i,i] \qquad \left(\text{sign}\left(\rho[i] \ominus \tau[i] - \frac{1}{2} \right) = \text{sign}\left(M^d_{\rho,\tau}[i,i] \right) \right) \\ &= \Omega(N^2) \sum_{i=1}^{K} |\alpha[i]|^2 \cdot d(\rho[i], \tau[i]) \qquad \text{(see (98) and (99))} \\ &\geq \Omega(N^2) \cdot \min_i |\alpha[i]|^2 \cdot \max_i d(\rho[i], \tau[i]) \\ &= \Omega(N^2) \cdot \min_i |\alpha[i]|^2 \cdot d(\rho, \tau) \qquad \text{(definition of Hausdorff distance in (8))} \\ &= \Omega(N^2) \cdot \frac{\min_i |\alpha[i]|^2}{\|\alpha\|_2^2} \cdot \|\alpha\|_2^2 \cdot d(\rho, \tau) \\ &= \Omega(K^{-1}N^2) \cdot \frac{\min_i |\alpha[i]|^2}{\max_i |\alpha[i]|^2} \cdot \|\alpha\|_2^2 \cdot d(\rho, \tau), \end{aligned}$$

and, consequently,

$$\begin{aligned}
&\left\langle M^d_{\rho,\tau}\alpha, M_{\rho,\tau}U\alpha\right\rangle \\
&\geq \frac{\Omega(K^{-1}N^2)}{\mathrm{dyn}(x_{\tau,\alpha})^2}\cdot\|\alpha\|_2^2\cdot d(\rho,\tau) \\
&= \frac{\Omega(N)}{\mathrm{dyn}(x_{\tau,\alpha})^2}\cdot\|\alpha\|_2^2\cdot d(\rho,\tau). \qquad \left(K\leq f_C+1=\frac{N+1}{2}\right)
\end{aligned}$$

With our choice of u earlier, we can substitute the bound above into (96) to finally obtain that

$$\begin{aligned}
\left\langle \frac{\partial F}{\partial\rho}(\rho), \mathrm{sign}\left((\rho\ominus\tau)-\frac{1}{2}\right)\right\rangle &= -\frac{\Omega(N)}{\mathrm{dyn}(x_{\tau,\alpha})^2}\cdot\|\alpha\|_2^2\cdot d(\rho,\tau)+O(e^{-Cc_2})\cdot\|\alpha\|_2^2 \\
&\quad + O(N)\cdot\|\widehat{n}_{\sigma_2}\|_2\,\|\alpha\|_2 + O(N)\cdot\|\widehat{n}_{\sigma_2}\|_2^2. \qquad (100)
\end{aligned}$$

The proof of Lemma 12 is complete because $\|\widehat{n}_{\sigma_2}\|_2\leq\|n(\cdot)\|_{L_2}$ (by (89)).

Supplementary Material

D Computing the Gradient of $F(\cdot)$

Here, for fixed $\rho_0\in\mathbb{I}^{\widetilde{K}}$, we wish to calculate $\frac{\partial F}{\partial\rho}(\rho_0)\in\mathbb{R}^{\widetilde{K}}$ and verify the explicit expression in (28). Set

$$\beta_{\rho_0}=\widetilde{\beta}\left(\rho_0\right):=G^{\dagger}_{\rho_0}\cdot\widehat{z}_{\sigma_2}\in\mathbb{R}^{\widetilde{K}}, \tag{101}$$

where, from (14), recall that the entries of $G_{\rho_0}\in\mathbb{C}^{N\times\widetilde{K}}$ are specified as

$$G_{\rho_0}[l,i]=\widehat{g}_{\sigma_2,N}[l]\cdot e^{-\mathrm{i}\,2\pi l\rho_0[i]}, \qquad l\in\mathbb{F},\ i\in[1:\widetilde{K}].$$

We use the following identity (which we later establish in Section F):

$$\frac{\partial F}{\partial\rho}(\rho_0)=\frac{\partial f}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right). \tag{102}$$

It suffices then to compute the right hand side of the above identity:

$$\begin{aligned}
\frac{\partial f}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) &= \left[\frac{\partial}{\partial\rho}\|G_\rho\beta-\widehat{z}_{\sigma_2}\|_2^2\right]\left(\rho_0,\widetilde{\beta}(\rho_0)\right) \\
&= \left[\frac{\partial}{\partial\rho}\left\langle G_\rho\beta-\widehat{z}_{\sigma_2}, G_\rho\beta-\widehat{z}_{\sigma_2}\right\rangle\right]\left(\rho_0,\widetilde{\beta}(\rho_0)\right) \\
&= 2\left(\frac{\partial G_\rho\beta}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\right)^*\left(G_{\rho_0}\cdot\widetilde{\beta}(\rho_0)-\widehat{z}_{\sigma_2}\right).
\end{aligned} \tag{103}$$

It only remains to calculate the derivative of $G_\rho\beta$ with respect to ρ. To that end, we next do some elementary calculations.

For $i\in[1:\widetilde{K}]$, we can compute the derivative of $G_\rho[:,i]\in\mathbb{C}^N$ (the ith column of $G_\rho\in\mathbb{C}^{N\times\widetilde{K}}$) with respect to $\rho[i]$ as

$$\frac{\partial\left(G_\rho[:,i]\right)}{\partial\rho[i]}\left(\rho_0[i]\right)=\begin{bmatrix}\vdots\\ \frac{\partial e^{-\mathrm{i}\,2\pi l\rho[i]}}{\partial\rho[i]}\left(\rho_0[i]\right)\\ \vdots\end{bmatrix}=L^*\cdot G_{\rho_0}[:,i]\in\mathbb{C}^N, \tag{104}$$

where the diagonal matrix $L \in \mathbb{C}^{N\times N}$ is specified by $L[l,l] = \mathrm{i}\,2\pi l$ for $l \in \mathbb{F}$. Above, for clarity, only the lth entry of the long vector is shown. In addition, for a vector $v \in \mathbb{R}^{\widetilde{K}}$, we observe that

$$
\begin{aligned}
\frac{\partial\left(G_\rho v\right)}{\partial\rho}(\rho_0) &= \sum_{i=1}^{\widetilde{K}} v[i]\cdot\left[\frac{\partial\left(G_\rho[:,i]\right)}{\partial\rho}\right](\rho_0) \\
&= \left[\begin{array}{ccc} \cdots & v[i]\cdot\frac{\partial G_\rho[:,i]}{\partial\rho[i]}\left(\rho_0[i]\right) & \cdots \end{array}\right] \in \mathbb{C}^{N\times\widetilde{K}} \\
&= \left[\begin{array}{ccc} \cdots & v[i]\cdot L^*\cdot G_{\rho_0}[:,i] & \cdots \end{array}\right] \qquad \text{(see (104))} \\
&= L^* G_{\rho_0}\cdot\operatorname{diag}(v)\,,
\end{aligned} \tag{105}
$$

where the second line follows because $G_\rho[:,i]$ depends only on $\rho[i]$. Above, $\operatorname{diag}(v) \in \mathbb{R}^{\widetilde{K}\times\widetilde{K}}$ is the diagonal vector formed from the entries of v. With (105) at hand, we can plug in for the derivitave of $G_\rho\beta$ in (103) to obtain that

$$
\begin{aligned}
&\frac{\partial F}{\partial\rho}(\rho_0) \\
&= \frac{\partial f}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) \\
&= 2\left(\frac{\partial G_\rho\beta}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\right)^*\left(G_{\rho_0}\cdot\widetilde{\beta}(\rho_0) - \widehat{z}_{\sigma_2}\right) \\
&= 2\left(L^* G_{\rho_0}\cdot\operatorname{diag}(\widetilde{\beta}(\rho_0))\right)^*\left(G_{\rho_0}\cdot\widetilde{\beta}(\rho_0) - \widehat{z}_{\sigma_2}\right) \\
&= 2\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)G_{\rho_0}^* L\left(G_{\rho_0}\cdot\widetilde{\beta}(\rho_0) - \widehat{z}_{\sigma_2}\right), \qquad \left(\widetilde{\beta}(\rho_0)\in\mathbb{R}^{\widetilde{K}}\right) \\
&= -2\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)G_{\rho_0}^* L\left(I_N - \mathcal{P}_{\rho_0}\right)\widehat{z}_{\sigma_2}, \qquad \text{(see (101))}
\end{aligned} \tag{106}
$$

where $\mathcal{P}_{\rho_0} = G_{\rho_0}G_{\rho_0}^\dagger$ is the orthogonal projection onto the column span of G_{ρ_0}. We therefore found an explicit expression for $\frac{\partial F}{\partial\rho}(\rho_0)$.

E Computing the Hessian of $F(\cdot)$

Here, for fixed $\rho_0 \in \mathbb{I}^{\widetilde{K}}$, we wish to calculate $\frac{\partial^2 F}{\partial\rho^2}(\rho_0) \in \mathbb{R}^{\widetilde{K}\times\widetilde{K}}$ and verify the explicit expression in (32). With $\widetilde{\beta}(\rho_0)$ as in (101), we will use the following identity (to be established in Section F):

$$
\begin{aligned}
\frac{\partial^2 F}{\partial\rho^2}(\rho_0) &= \frac{\partial^2 f}{\partial\rho^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) + 2\cdot\frac{\partial^2 f}{\partial\rho\partial\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0) \\
&\quad + \left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0)\right)^*\cdot\frac{\partial^2 f}{\partial\beta^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0).
\end{aligned} \tag{107}
$$

We are now burdened with the laborious task of computing the following derivatives:

$$
\frac{\partial^2 f}{\partial\rho^2}(\rho,\beta)\,, \qquad \frac{\partial^2 f}{\partial\rho\partial\beta}(\rho,\beta)\,, \qquad \frac{\partial^2 f}{\partial\beta^2}(\rho,\beta)\,, \qquad \frac{\partial\widetilde{\beta}}{\partial\rho}(\rho). \tag{108}
$$

Recall (104) and (105) to facilitate the ensuing arguments. Three fresh estimates are needed before calculating the derivatives in (108). These estimates will be presented immediately next and then followed by the body of calculations throughout the rest of this section. As for the first auxiliary result, for a vector $u \in \mathbb{R}^N$, we note that

$$
\begin{aligned}
\frac{\partial\left(G_\rho^* u\right)}{\partial \rho}(\rho) &= \begin{bmatrix} \vdots \\ \frac{\partial((G_\rho[:,i])^* u)}{\partial \rho}(\rho) \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & & \\ & \frac{\partial((G_\rho[:,i])^* u)}{\partial \rho[i]}(\rho[i]) & \\ & & \ddots \end{bmatrix} \\
&= \begin{bmatrix} \ddots & & \\ & \left(\frac{\partial(G_\rho[:,i])}{\partial \rho[i]}(\rho[i])\right)^* u & \\ & & \ddots \end{bmatrix} \\
&= \begin{bmatrix} \ddots & & \\ & (G_\rho[:,i])^* Lu & \\ & & \ddots \end{bmatrix} \qquad \text{(see (104))} \\
&= \operatorname{diag}\left(G_\rho^* L u\right) \in \mathbb{R}^{\widetilde{K}\times\widetilde{K}}, \qquad (109)
\end{aligned}
$$

where the second identity holds because $G_\rho[:,i]$ depends only on $\rho[i]$. Also, note that

$$
\begin{aligned}
\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni \frac{\partial\left(G_\rho^* G_\rho \beta\right)}{\partial \rho}(\rho) &= \begin{bmatrix} \vdots \\ \frac{\partial((G_\rho[:,i])^* G_\rho \beta)}{\partial \rho}(\rho) \\ \vdots \end{bmatrix} \\
&= \begin{bmatrix} \vdots \\ \left(\frac{\partial(G_\rho[:,i])}{\partial \rho}(\rho)\right)^* G_\rho \beta \\ \vdots \end{bmatrix} + \begin{bmatrix} \vdots \\ (G_\rho[:,i])^* \cdot \frac{\partial(G_\rho \beta)}{\partial \rho}(\rho) \\ \vdots \end{bmatrix} \\
&= \begin{bmatrix} \ddots & & \\ & (L^* \cdot G_\rho[:,i])^* G_\rho \beta & \\ & & \ddots \end{bmatrix} \\
&\qquad + \begin{bmatrix} \vdots \\ (G_\rho[:,i])^* L^* G_\rho \cdot \operatorname{diag}(\beta) \\ \vdots \end{bmatrix} \qquad \text{(see (104) and (105))} \\
&= \begin{bmatrix} \ddots & & \\ & (G_\rho[:,i])^* L G_\rho \beta & \\ & & \ddots \end{bmatrix} + \begin{bmatrix} \vdots \\ (G_\rho[:,i])^* L^* G_\rho \cdot \operatorname{diag}(\beta) \\ \vdots \end{bmatrix},
\end{aligned}
$$

and, consequently,

$$
\begin{aligned}
\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni & \frac{\partial\left(G_\rho^* G_\rho \beta\right)}{\partial \rho}(\rho) \\
&= \operatorname{diag}\left(G_\rho^* L G_\rho \beta\right) + G_\rho^* L^* G_\rho \cdot \operatorname{diag}(\beta). \qquad (110)
\end{aligned}
$$

Similarly,

$$\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni \frac{\partial\left(G_\rho^* L G_\rho \beta\right)}{\partial\rho}(\rho,\beta) = \operatorname{diag}\left(G_\rho^* L^2 G_\rho \beta\right) + G_\rho^* L L^* G_\rho \cdot \operatorname{diag}(\beta)$$
$$= \operatorname{diag}\left(G_\rho^* L^2 G_\rho \beta\right) - G_\rho^* L^2 G_\rho \cdot \operatorname{diag}(\beta). \qquad (L^* = -L) \tag{111}$$

Armed with the necessary estimates, we embark on calculating the derivatives in (108). Beginning with $\frac{\partial^2 f}{\partial \rho^2}(\cdot,\cdot)$, note that

$$\begin{aligned}\mathbb{R}^{\widetilde{K}\times\widetilde{K}} &\ni \frac{\partial^2 f}{\partial\rho^2}(\rho,\beta)\\ &= \frac{\partial}{\partial\rho}\left(\frac{\partial f}{\partial\rho}(\rho,\beta)\right)\\ &= \left[\frac{\partial}{\partial\rho}\left(2\cdot\operatorname{diag}(\beta)\, G_\rho^* L\left(G_\rho\beta - \widehat{z}_{\sigma_2}\right)\right)\right](\rho,\beta) \qquad \text{(see (106))},\end{aligned}$$

and, consequently,

$$\begin{aligned}\mathbb{R}^{\widetilde{K}\times\widetilde{K}} &\ni \frac{\partial^2 f}{\partial\rho^2}(\rho,\beta)\\ &= 2\cdot\operatorname{diag}(\beta)\cdot\frac{\partial\left(G_\rho^* L\left(G_\rho\beta-\widehat{z}_{\sigma_2}\right)\right)}{\partial\rho}(\rho,\beta)\\ &= 2\cdot\operatorname{diag}(\beta)\cdot\frac{\partial\left(G_\rho^* L G_\rho\beta\right)}{\partial\rho}(\rho,\beta) - 2\cdot\operatorname{diag}(\beta)\cdot\frac{\partial\left(G_\rho^* L\widehat{z}_{\sigma_2}\right)}{\partial\rho}(\rho,\beta)\\ &= -2\cdot\operatorname{diag}(\beta)\cdot G_\rho^* L^2 G_\rho\cdot\operatorname{diag}(\beta) + 2\cdot\operatorname{diag}(\beta)\cdot\operatorname{diag}\left(G_\rho^* L^2 G_\rho\beta\right)\\ &\qquad - 2\cdot\operatorname{diag}(\beta)\cdot\operatorname{diag}\left(G_\rho^* L^2\widehat{z}_{\sigma_2}\right) \qquad \text{(see (111) and (109))}\\ &= -2\cdot\operatorname{diag}(\beta)\cdot G_\rho^* L^2 G_\rho\cdot\operatorname{diag}(\beta)\\ &\qquad + 2\cdot\operatorname{diag}(\beta)\cdot\operatorname{diag}\left(G_\rho^* L^2\left(G_\rho\beta-\widehat{z}_{\sigma_2}\right)\right).\end{aligned}$$

In particular, using (101), we find that

$$\frac{\partial^2 f}{\partial\rho^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) = -2\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)\cdot G_{\rho_0}^* L^2 G_{\rho_0}\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)$$
$$- 2\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)\cdot\operatorname{diag}\left(G_{\rho_0}^* L^2\left(I_N-\mathcal{P}_{\rho_0}\right)\widehat{z}_{\sigma_2}\right). \tag{112}$$

As usual, $\mathcal{P}_{\rho_0} = G_{\rho_0} G_{\rho_0}^\dagger$. In a similar fashion, we compute $\frac{\partial^2 f}{\partial\beta\partial\rho}(\cdot,\cdot)$ by writing that

$$\begin{aligned}
\mathbb{R}^{\widetilde{K}\times\widetilde{K}} \ni & \frac{\partial^2 f}{\partial\beta\partial\rho}(\rho,\beta) \\
&= \frac{\partial}{\partial\rho}\left(\frac{\partial f}{\partial\beta}(\rho,\beta)\right) \\
&= \frac{\partial}{\partial\rho}\left(\frac{\partial\left\|G_\rho\beta - \widehat{z}_{\sigma_2}\right\|_2^2}{\partial\beta}(\rho,\beta)\right) \qquad \text{(see (21))} \\
&= \left[\frac{\partial}{\partial\rho}\left(2\cdot G_\rho^*\left(G_\rho\beta - \widehat{z}_{\sigma_2}\right)\right)\right](\rho,\beta) \\
&= 2\cdot\frac{\partial\left(G_\rho^* G_\rho\beta\right)}{\partial\rho}(\rho,\beta) - \frac{\partial\left(G_\rho^*\widehat{z}_{\sigma_2}\right)}{\partial\rho}(\rho,\beta) \\
&= 2\cdot G_\rho^* L^* G_\rho\cdot\operatorname{diag}(\beta) + 2\cdot\operatorname{diag}\left(G_\rho^* L G_\rho\beta\right) \\
&\qquad - 2\cdot\operatorname{diag}\left(G_\rho^* L\widehat{z}_{\sigma_2}\right) \qquad \text{(see (109) and(110))} \\
&= 2\cdot G_\rho^* L^* G_\rho\cdot\operatorname{diag}(\beta) + 2\cdot\operatorname{diag}\left(G_\rho^* L\left(G_\rho\beta - \widehat{z}_{\sigma_2}\right)\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{\partial^2 f}{\partial\rho\partial\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) \\
&= \left(\frac{\partial^2 f}{\partial\beta\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\right)^* \\
&= 2\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)\cdot G_{\rho_0}^* L^* G_{\rho_0} + 2\cdot\operatorname{diag}\left(G_{\rho_0}^* L\left(G_{\rho_0}\widetilde{\beta}(\rho_0) - \widehat{z}_{\sigma_2}\right)\right) \\
&= 2\cdot\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right)\cdot G_{\rho_0}^* L^* G_{\rho_0} - 2\cdot\operatorname{diag}\left(G_{\rho_0}^* L\left(I_N - \mathcal{P}_{\rho_0}\right)\widehat{z}_{\sigma_2}\right).
\end{aligned} \tag{113}$$

Also,

$$\begin{aligned}
\frac{\partial^2 f}{\partial\beta^2}(\rho,\beta) &= \frac{\partial}{\partial\beta}\left(\frac{\partial\left\|G_\rho\beta - \widehat{z}_{\sigma_2}\right\|_2^2}{\partial\beta}(\rho,\beta)\right) \qquad \text{(see (21))} \\
&= 2\cdot\frac{\partial\left(G_\rho^*\left(G_\rho\beta - \widehat{z}_{\sigma_2}\right)\right)}{\partial\beta}(\rho,\beta) \\
&= 2G_\rho^* G_\rho,
\end{aligned}$$

and, clearly,

$$\frac{\partial^2 f}{\partial\beta^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) = 2G_{\rho_0}^* G_{\rho_0}. \tag{114}$$

Lastly, in order to compute $\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)$, recall from (101) that

$$\widetilde{\beta}(\rho) = G_\rho^\dagger\widehat{z}_{\sigma_2} = \left(G_\rho^* G_\rho\right)^{-1} G_\rho^*\widehat{z}_{\sigma_2},$$

or, equivalently,

$$G_\rho^* G_\rho\cdot\widetilde{\beta}(\rho) = G_\rho^*\widehat{z}_{\sigma_2}.$$

The ith row of the above identity reads

$$\left(G_\rho^*\cdot G_\rho[:,i]\right)^*\widetilde{\beta}(\rho) = \left(G_\rho[:,i]\right)^*\widehat{z}_{\sigma_2}.$$

Taking derivatives of both sides (with respect to ρ) yields

$$\left(\frac{\partial\left(G_\rho^*\cdot G_\rho[:,i]\right)}{\partial\rho}(\rho)\right)^*\widetilde{\beta}(\rho)+\left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)\right)^*G_\rho^T\cdot\overline{G_\rho[:,i]}=\left(\frac{\partial G_\rho[:,i]}{\partial\rho}(\rho)\right)^*\widehat{z}_{\sigma_2},$$

where a^T is the transpose of vector a, and $\overline{b}$ denotes the complex conjugate of scalar b. After rearranging to isolate the target term $\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)$, we continue to simplify the above identity:

$$\begin{aligned}
&\left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)\right)^*G_\rho^T\cdot\overline{G_\rho[:,i]}\\
&=\left(\frac{\partial G_\rho[:,i]}{\partial\rho}(\rho)\right)^*\widehat{z}_{\sigma_2}-\left(\frac{\partial\left(G_\rho^*\cdot G_\rho[:,i]\right)}{\partial\rho}(\rho)\right)^*\widetilde{\beta}(\rho)\\
&=\left(\frac{\partial G_\rho[:,i]}{\partial\rho}(\rho)\right)^*\widehat{z}_{\sigma_2}\\
&\qquad-\left(\frac{\partial\left(G_\rho^*G_\rho\cdot e_i\right)}{\partial\rho}(\rho)\right)^*\widetilde{\beta}(\rho),\qquad\left(e_i:\ i\text{th canonical vector in }\mathbb{R}^{\widetilde{K}}\right)
\end{aligned}$$

and, consequently,

$$\begin{aligned}
&\left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)\right)^*G_\rho^T\cdot\overline{G_\rho[:,i]}\\
&=\left(\left(L^*\cdot G_\rho[:,i]\right)^*\widehat{z}_{\sigma_2}\right)\cdot e_i\\
&\qquad-\left(G_\rho^*L^*G_\rho\cdot\operatorname{diag}(e_i)+\operatorname{diag}\left(G_\rho^*LG_\rho e_i\right)\right)^*\widetilde{\beta}(\rho)\quad\text{(see (104) and (110))}\\
&=\left(\left(L^*\cdot G_\rho[:,i]\right)^*\widehat{z}_{\sigma_2}\right)\cdot e_i-\operatorname{diag}(e_i)\cdot G_\rho^*LG_\rho\cdot\widetilde{\beta}(\rho)-\operatorname{diag}\left(e_i^*G_\rho^*L^*G_\rho\right)\cdot\widetilde{\beta}(\rho)\\
&=\left(\left(L^*\cdot G_\rho[:,i]\right)^*\widehat{z}_{\sigma_2}\right)\cdot e_i-\left(\left(G_\rho[:,i]\right)^*LG_\rho\cdot\widetilde{\beta}(\rho)\right)\cdot e_i-\operatorname{diag}\left(\left(G_\rho[:,i]\right)^*L^*G_\rho\right)\cdot\widetilde{\beta}(\rho)\\
&=\left(\left(G_\rho[:,i]\right)^*L\widehat{z}_{\sigma_2}\right)\cdot e_i-\left(\left(G_\rho[:,i]\right)^*LG_\rho\cdot\widetilde{\beta}(\rho)\right)\cdot e_i-\operatorname{diag}(\widetilde{\beta}(\rho))\cdot G_\rho^*LG_\rho[:,i]\\
&=\left(\left(G_\rho[:,i]\right)^*L\left(\widehat{z}_{\sigma_2}-G_\rho\cdot\widetilde{\beta}(\rho)\right)\right)\cdot e_i-\operatorname{diag}(\widetilde{\beta}(\rho))\cdot G_\rho^*LG_\rho[:,i].
\end{aligned}\tag{115}$$

The second to last line above uses the identity $\operatorname{diag}(a)\cdot b=\operatorname{diag}(b)\cdot a$ for vectors a and b of the same length. By stacking the columns for all values of i, we obtain that

$$\begin{aligned}
&\left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)\right)^*G_\rho^T\overline{G_\rho}\\
&=\begin{bmatrix}\ddots&&\\&\left(G_\rho[:,i]\right)^*L\left(\widehat{z}_{\sigma_2}-G_\rho\cdot\widetilde{\beta}(\rho)\right)&\\&&\ddots\end{bmatrix}-\operatorname{diag}(\widetilde{\beta}(\rho))\cdot G_\rho^*LG_\rho\qquad\text{(see (115))}\\
&=\operatorname{diag}\left(G_\rho^*L\left(\widehat{z}_{\sigma_2}-G_\rho\cdot\widetilde{\beta}(\rho)\right)\right)-\operatorname{diag}(\widetilde{\beta}(\rho))\cdot G_\rho^*LG_\rho,
\end{aligned}$$

or

$$G_\rho^*G_\rho\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho)=\operatorname{diag}\left(G_\rho^*L\left(\widehat{z}_{\sigma_2}-G_\rho\cdot\widetilde{\beta}(\rho)\right)\right)-G_\rho^*L^*G_\rho\cdot\operatorname{diag}(\widetilde{\beta}(\rho)).$$

We conclude that

$$
\begin{aligned}
\frac{\partial \widetilde{\beta}}{\partial \rho}(\rho) &= \left(G_\rho^* G_\rho\right)^{-1} \operatorname{diag}\left(G_\rho^* L \left(\widehat{z}_{\sigma_2} - G_\rho \cdot \widetilde{\beta}(\rho)\right)\right) - \left(G_\rho^* G_\rho\right)^{-1} G_\rho^* L^* G_\rho \cdot \operatorname{diag}(\widetilde{\beta}(\rho)) \\
&= \left(G_\rho^* G_\rho\right)^{-1} \operatorname{diag}\left(G_\rho^* L \left(\widehat{z}_{\sigma_2} - G_\rho \cdot \widetilde{\beta}(\rho)\right)\right) - G_\rho^\dagger L^* G_\rho \cdot \operatorname{diag}(\widetilde{\beta}(\rho)),
\end{aligned}
$$

and, in particular,

$$
\frac{\partial \widetilde{\beta}}{\partial \rho}(\rho_0) = \left(G_{\rho_0}^* G_{\rho_0}\right)^{-1} \operatorname{diag}\left(G_{\rho_0}^* L \left(I_N - \mathcal{P}_{\rho_0}\right) \widehat{z}_{\sigma_2}\right) - G_{\rho_0}^\dagger L^* G_{\rho_0} \cdot \operatorname{diag}(\widetilde{\beta}(\rho_0)). \tag{116}
$$

To summarize, we finished computing all the quantities involved in (107) (see (112-114), and (116)). We can simplify the above expression for the Hessian of $F(\cdot)$ by noting that the second and third summands in (107) differ only by a constant factor. More specifically, from (113) and (116), it follows that

$$
\begin{aligned}
&\frac{\partial^2 f}{\partial \rho \partial \beta}\left(\rho_0, \widetilde{\beta}(\rho_0)\right) \cdot \frac{\partial \widetilde{\beta}}{\partial \rho}(\rho_0) \\
&= -2\left[\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right) G_{\rho_0}^* L G_{\rho_0} - \operatorname{diag}\left(G_{\rho_0}^* L \left(I_N - \mathcal{P}_{\rho_0}\right) \widehat{z}_{\sigma_2}\right)\right] \\
&\qquad \cdot \left(G_{\rho_0}^* G_{\rho_0}\right)^{-1} \cdot \left[G_{\rho_0}^* L^* G_{\rho_0} \cdot \operatorname{diag}(\widetilde{\beta}(\rho_0)) - \operatorname{diag}\left(G_{\rho_0}^* L \left(I_N - \mathcal{P}_{\rho_0}\right) \widehat{z}_{\sigma_2}\right)\right], \\
&= -\left(\frac{\partial \widetilde{\beta}}{\partial \rho}(\rho_0)\right)^* \cdot \frac{\partial^2 f}{\partial \beta^2}\left(\rho_0, \widetilde{\beta}(\rho_0)\right) \cdot \frac{\partial \widetilde{\beta}}{\partial \rho}(\rho_0),
\end{aligned}
$$

so that

$$
\begin{aligned}
\frac{\partial^2 F}{\partial \rho^2}(\rho_0) &= \frac{\partial^2 f}{\partial \rho^2}\left(\rho_0, \widetilde{\beta}(\rho_0)\right) + \frac{\partial^2 f}{\partial \rho \partial \beta}\left(\rho_0, \widetilde{\beta}(\rho_0)\right) \cdot \frac{\partial \widetilde{\beta}}{\partial \rho}(\rho_0) \\
&= -2 \cdot \operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right) \cdot G_{\rho_0}^* L^2 G_{\rho_0} \cdot \operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right) \\
&\qquad - 2 \cdot \operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right) \cdot \operatorname{diag}\left(G_{\rho_0}^* L^2 \left(I_N - \mathcal{P}_{\rho_0}\right) \widehat{z}_{\sigma_2}\right) \\
&\qquad - 2\left[\operatorname{diag}\left(\widetilde{\beta}(\rho_0)\right) G_{\rho_0}^* L G_{\rho_0} - \operatorname{diag}\left(G_{\rho_0}^* L \left(I_N - \mathcal{P}_{\rho_0}\right) \widehat{z}_{\sigma_2}\right)\right] \\
&\qquad \cdot \left(G_{\rho_0}^* G_{\rho_0}\right)^{-1} \cdot \left[G_{\rho_0}^* L^* G_{\rho_0} \cdot \operatorname{diag}(\widetilde{\beta}(\rho_0)) - \operatorname{diag}\left(G_{\rho_0}^* L \left(I_N - \mathcal{P}_{\rho_0}\right) \widehat{z}_{\sigma_2}\right)\right.,
\end{aligned} \tag{117}
$$

which might be simplified slightly further.

F Ingredients for Computing $\frac{\partial F}{\partial \rho}(\cdot)$ and $\frac{\partial^2 F}{\partial \rho^2}(\cdot)$

Here, we establish (102) and (107). Fix ρ_0 and suppose that $f(\cdot, \cdot)$ is analytic, i.e., has convergent power series everywhere . Moreover, assume that

$$
\widetilde{\beta}(\rho) := \arg\min_{\beta} f(\rho, \beta)
$$

is always well-defined, i.e., $\widetilde{\beta}(\rho)$ is the unique minimizer of $f(\rho, \cdot)$ for every ρ. In particular, by implicit function theorem, $\widetilde{\beta}(\rho)$ is smooth (i.e., infinitely differentiable with respect to ρ). We wish to calculate the first and second derivatives of $F(\cdot)$, the map that takes ρ to $F(\rho) = \min_\beta f(\rho,\beta) = f(\rho,\widetilde{\beta}(\rho))$. (The existence of these derivatives is established along the way.)

To that end, we note that the following expansion holds for small enough $|\rho-\rho_0|$ and $|\widetilde{\beta}(\rho) - \widetilde{\beta}(\rho_0)|$:

$$\begin{aligned}
F(\rho) &= f\left(\rho,\widetilde{\beta}(\rho)\right) \\
&= f\left(\rho_0,\widetilde{\beta}(\rho_0)\right) + (\rho-\rho_0)^T\cdot\frac{\partial f}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) + \left(\widetilde{\beta}(\rho)-\widetilde{\beta}(\rho_0)\right)^T\cdot\frac{\partial f}{\partial\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) \\
&\quad + \frac{1}{2}(\rho-\rho_0)^T\cdot\frac{\partial^2 f}{\partial\rho^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot(\rho-\rho_0) \\
&\quad + (\rho-\rho_0)^T\cdot\frac{\partial^2 f}{\partial\rho\partial\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\left(\widetilde{\beta}(\rho)-\widetilde{\beta}(\rho_0)\right) \\
&\quad + \frac{1}{2}\left(\widetilde{\beta}(\rho)-\widetilde{\beta}(\rho_0)\right)^T\cdot\frac{\partial^2 f}{\partial^2\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\left(\widetilde{\beta}(\rho)-\widetilde{\beta}(\rho_0)\right) + o_2.
\end{aligned}$$

Above, o_2 comprises of negligible terms. Note that $f(\rho_0,\widetilde{\beta}(\rho_0)) = F(\rho_0)$ and that

$$\frac{\partial f}{\partial\beta}(\rho_0,\widetilde{\beta}(\rho_0)) = 0,$$

because $\widetilde{\beta}(\rho_0)$ minimizes $f(\rho_0,\cdot)$. On the other hand, because $\widetilde{\beta}(\rho)$ is a smooth function of ρ, $\widetilde{\beta}(\rho)-\widetilde{\beta}(\rho_0) = \frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0)\cdot(\rho-\rho_0)+o_1$ for small enough $|\rho-\rho_0|$. Here, o_1 collects the negligible terms.Therefore, the above expansion simplifies to

$$\begin{aligned}
F(\rho) = F(\rho_0) &+ (\rho-\rho_0)^T\cdot\frac{\partial f}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) + \frac{1}{2}(\rho-\rho_0)^T\cdot\frac{\partial^2 f}{\partial\rho^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot(\rho-\rho_0) \\
&+ (\rho-\rho_0)^T\cdot\frac{\partial^2 f}{\partial\rho\partial\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0)\cdot(\rho-\rho_0) \\
&+ \frac{1}{2}(\rho-\rho_0)^T\cdot\left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0)\right)^T\cdot\frac{\partial^2 f}{\partial\beta^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0)\cdot(\rho-\rho_0) + o_2.
\end{aligned}$$

We conclude that

$$\frac{\partial F}{\partial\rho}(\rho_0) = \frac{\partial f}{\partial\rho}\left(\rho_0,\widetilde{\beta}(\rho_0)\right),$$

$$\begin{aligned}
\frac{\partial^2 F}{\partial\rho^2}(\rho_0) = \frac{\partial^2 f}{\partial\rho^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right) &+ 2\cdot\frac{\partial^2 f}{\partial\rho\partial\beta}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0) \\
&+ \left(\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0)\right)^T\cdot\frac{\partial^2 f}{\partial\beta^2}\left(\rho_0,\widetilde{\beta}(\rho_0)\right)\cdot\frac{\partial\widetilde{\beta}}{\partial\rho}(\rho_0).
\end{aligned}$$

Note that, despite the nonsymmetric appearance of the second term in the Hessian, $\frac{d^2F}{d\rho^2}(\rho_0) \in \mathbb{R}^{K\times K}$ is indeed a symmetric matrix.